



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Young, JT;van Dooren, K;Lennox, NG;Butler, TG;Kinner, SA

Title:

Inter-rater reliability of the Hayes Ability Screening Index in a sample of Australian prisoners

Date:

2015-11-01

Citation:

Young, J. T., van Dooren, K., Lennox, N. G., Butler, T. G. & Kinner, S. A. (2015). Inter-rater reliability of the Hayes Ability Screening Index in a sample of Australian prisoners. *Journal of Intellectual Disability Research*, 59 (11), pp.1055-1060. <https://doi.org/10.1111/jir.12198>.

Persistent Link:

<https://hdl.handle.net/11343/57293>

Inter-rater reliability of the Hayes Ability Screening Index in a sample of Australian prisoners

Jesse T Young^{1,2,3} (MPH), T: +61 3 8344 4413, F: +61 3 9348 1174, jesse.young@unimelb.edu.au;

Kate van Dooren⁴ (PhD), T: +61 7 3163 2412, F: +61 7 3163 2445, k.vandooren@sph.uq.edu.au;

Nick G Lennox⁴ (FRACGP PhD), T: +61 7 3163 2413, F: +61 7 3163 2445, n.lennox@uq.edu.au;

Tony G Butler⁵ (PhD), T: +61 2 9289 2928, F: +61 2 9385 0920, tbutler@kirby.unsw.edu.au;

Stuart A Kinner^{1,6,7,8} (PhD), T: +61 3 9035 7598, F: +61 3 9348 1174, s.kinner@unimelb.edu.au

¹Melbourne School of Population and Global Health, The University of Melbourne, Carlton, VIC, 3010, Australia

²Centre for Health Services Research, School of Population Health, The University of Western Australia, Perth, WA, 6009, Australia

³National Drug Research Institute, Curtin University, Perth, WA, 6008, Australia

⁴Queensland Centre for Intellectual and Developmental Disability, School of Medicine, The University of Queensland, Mater Misericordiae Hospital, South Brisbane, QLD, 4101, Australia

⁵Kirby Institute, UNSW Australia, Sydney, NSW, 2052, Australia

⁶School of Medicine, The University of Queensland, Herston, QLD, 4006, Australia

⁷School of Public Health and Preventive Medicine, Monash University, The Alfred Centre, Melbourne, VIC, 3004, Australia

⁸Murdoch Children's Research Institute, Parkville, VIC, 3052, Australia

Word count (Introduction – Conclusion): 1498; (Abstract): 230; Number of pages: 11; Number of tables: 2

Corresponding Author:

Mr Jesse T Young

The University of Melbourne

Level 4, 207 Bouverie Street, Carlton, VIC, 3010, Australia

Tel: +61 3 8344 4413; Fax: +61 3 9348 1174

Email: jesse.young@unimelb.edu.au

Note: This is the pre-peer reviewed version of the following article: Young J T., van Dooren K., Lennox N G., Butler T G., Kinner S A. Inter-rater reliability of the Hayes Ability Screening Index in a sample of Australian prisoners. *Journal of Intellectual Disability Research*. 2015; *Epub ahead of print.*, which has been published in final form at <http://dx.doi.org/10.1111/jir.12198>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

ABSTRACT

Background:

Reliable ascertainment of intellectual disability (ID) is important to identify those with special needs, in order for those needs to be met in the criminal justice system. Although the Hayes Ability Screening Index (HASI) is valid and widely used for the identification of possible ID, the risk of inter-rater bias between researchers when scoring the HASI has not yet been established. The current paper estimates the inter-rater reliability of the HASI in a sample of Indigenous and non-Indigenous prisoners in Western Australia.

Methods:

We estimated intra-class correlation coefficients (ICC) for the consistency of agreement among three blinded raters using a two-way random effects model assessing the inter-rater agreement of the HASI. Kappa was also estimated for the dichotomous HASI screening threshold outcome between the raters.

Results:

The HASI exhibited very good within-subject consistency of agreement for Section B (ICC=0.95; 95%CI:0.94-0.96), Section C (ICC=0.97; 95%CI:0.96-0.98), and Section D (ICC=0.90; 95%CI:0.87-0.92) subscales as well as for the total scaled score (ICC=0.97; 95%CI:0.96-0.98). The inter-rater reliability of the dichotomous adult ID screening threshold (<85) was also very good (Kappa=0.95).

Conclusions:

The current study provides new evidence that the HASI has a low risk of bias from between-rater scoring and can be reliably scored by both non-clinicians and clinicians with little training, when administered in prison settings. Pre-scoring training should focus on the more subjective 'clock-drawing' section, in order to maximise inter-rater reliability.

Key Words: Intellectual Disability; Reliability; Screening; Prisoners; Scoring Methods

INTRODUCTION

People with intellectual disability (ID) are over-represented in prisons internationally (Baldry *et al.* 2013); however, ascertainment of ID on reception is difficult and resource-intensive (Hayes *et al.* 2007) and not usual practice in most prison settings (Simpson *et al.* 2001). Reliable screening is important to identify those with specific needs not identified prior to incarceration, given evidence of under-utilisation of clinical and disability services in offender populations pre-incarceration (Baldry *et al.* 2012, AIHW 2013, Indig *et al.* 2010).

The Hayes Ability Screening Index (HASI) is a screening tool for the identification of possible ID, designed for administration by non-clinicians (Hayes 2000). It consists of four sections: A) a Background section comprising four yes/no questions assessing a likely history of learning problems; B) a Backwards-spelling section where participants are asked to spell a five-letter word in reverse-order; C) a Puzzle section requiring participants to join circles with numbers and letters in an alternating, ascending sequence; and D) a Clock-drawing section necessitating drawing a large clock-face with the clock-hands at a specified time (Hayes 2000).

The HASI displays adequate construct and external validity among prisoners (Hayes 2000, Hayes 2002), however due to being over-inclusive of borderline ID, it has been shown to yield some false-positives when compared to the Wechsler Adult Intelligence Scale in prisoners (Butler and Milner 2003). Nevertheless, inconsistencies between those scoring the screening observations (i.e., raters) can be a source of information bias, potentially impacting the reliability of screening results and the interpretation of data collected. If the HASI is to be applied in research and practice, where it is typically scored by different team members, it is important to establish its inter-rater reliability (IRR).

The aim of the current study was to estimate the IRR of the HASI, when scored blindly by different raters with basic training, in a sample of prisoners in Western Australia (WA).

METHODS

Participants

A structured interview was administered to 190 adult prisoners (≥ 18 years) within six-weeks of their expected release from two prisons in Perth, WA between 28 May 2013 and 22 May 2014. All participants provided informed written consent.

Assessments

Baseline self-report measures included age, gender, Indigenous status, country of origin, usual accommodation before prison, and pre-incarceration employment status. Three trained research interviewers administered the HASI for the identification of possible ID (Hayes 2000) in one male and one female prison in Perth, WA. The HASI produces a total scaled score (range: 48.7-96.4) with scores < 85 considered consistent with possible adult ID (Hayes 2000). This total score is comprised of four section scores (A: Background; B: Backwards-spelling test; C: Puzzle; D: Clock-drawing). A full description of the HASI scoring methodology is provided elsewhere (Hayes 2000).

The HASI for each participant was scored separately by three raters who were blinded to participant characteristics and each other's score. Raters one and two were experienced researchers who received HASI-specific training. Rater three was a registered psychologist with extensive experience in ID screening and clinical assessment. Prior to scoring the HASI, all three raters participated in a teleconference to establish scoring criteria.

Statistical Analysis

Analyses were conducted using STATA v13.1 (StataCorp 2013). Descriptive statistics were calculated for all variables. Scaled scores were calculated for each section of the HASI and summed into a total scaled score as outlined elsewhere (Hayes 2000). Intra-class correlation coefficients (ICC) for the individual and average consistency of agreement among the three raters were estimated using a two-way random-effects model (Shrout and Fleiss 1979). The

ICC is the recommended measure for assessing IRR on continuous measures as it takes into account rating variability and any systematic differences between the raters (Rousson *et al.* 2002). An ICC was estimated separately for each HASI section and for the total scaled score. Differences in consistency of agreement across gender and Indigenous status were compared using post-estimation F-tests.

In accordance with standard practice, the total scaled score was dichotomised for each rater, with scores <85 indicating possible adult ID (Hayes 2000). A Kappa statistic was calculated for the agreement among all three raters.

Ethical Considerations

The current study was approved by the University of Western Australia Human Research Ethics Committee and the WA Department of Corrective Services Research and Evaluation Committee.

RESULTS

Participant characteristics are described in Table 1. The sample was predominantly male (n=170;89.5%). Almost half of participants identified as Indigenous Australian (n=89;46.8%). The mean age (\pm standard deviation) was 29.9 (\pm 9.1) years with more than a third aged from 18 to 24 years (Table 1).

HASI scaled score descriptive statistics and ICC estimates for IRR are presented in Table 2. Collapsed across raters, the average HASI screening score was 85.6 (\pm 10.1; range: 58.2-96.4). The ICC estimates indicated very good within-subject consistency of agreement for section B (ICC=0.95; 95%CI:0.94-0.96), section C (ICC=0.97; 95%CI:0.96-0.98), and section D (ICC=0.90; 95%CI:0.87-0.92) subscales as well as for the total scaled score (ICC=0.97; 95%CI:0.96-0.98) (Table 2). Section A is an objective dichotomous measure; we observed perfect concordance indicating no data-entry inconsistencies.

Subgroup analysis indicated greater within-subject agreement when scoring Indigenous participants compared to their non-Indigenous counterparts($F=4.00;p<0.001$). The ICC estimates did not significantly differ between males and females($F=0.69;p=0.809$). The ICC estimates for the average agreement between raters (i.e., the correlation of mean ratings for all observations) was higher for every measure as compared to individual-level rater-agreement (i.e., the correlation of ratings made on the same participant). The IRR of the dichotomous adult ID screening threshold was very good($Kappa=0.95$)(Table 2).

Table 1. Sociodemographic characteristics of study participants

Participant Characteristics	All participants N = 190	Percentage
Gender		
- Male	170	89.5%
Age		
-18-24 Years	68	35.8%
-25-29 Years	42	22.1%
-30-39 Years	52	27.4%
-≥40 years	28	14.7%
Indigenous Status		
-Yes	89	46.8%
Employment Status		
-Full-time work	57	30.0%
-Part-time work	20	10.5%
-Student/Home Duties/ Pension	19	10.0%
-Unemployed / Criminal Activity	94	49.5%
Usual Accommodation		
- Owned House / Flat	9	4.7%
- Rented House / Flat	71	37.4%
- Parent's Home	35	18.4%
- Public Housing	29	15.3%
- Temporary Accommodation	24	12.6%
- Shelter / NFA / Homeless	15	7.9%
- Other	7	3.7%
Country of Origin ^a		
- Australia	169	89.4%

^aTotal sums to less than 100% due to missing data (n=1; 0.5%)

NFA: No Fixed Address

RUNNING HEAD: HASI Inter-rater Reliability in Prisoners

Table 2. HASI Inter-rater reliability

HASI Scaled Score (N=190)	Rater 1	Rater 2	Rater 3	Total N=570	ICC(95%CI) Marginal	Subject
Section Scores						
Section A						
- Range (Min, Max)	14(14, 28)	14(14, 28)	14(14, 28)	14(14, 28)		
- Mean(\pm SD)	25.97(\pm 3.55)	25.97(\pm 3.55)	25.97(\pm 3.55)	25.97(\pm 3.55)	1(-)1	1(-)1
Section B						
- Range (Min, Max)	12(3, 15)	12(3, 15)	12(3, 15)	12(3, 15)		
- Mean(\pm SD)	12.44(\pm 4.61)	12.24(\pm 4.80)	12.54(\pm 4.50)	12.41(\pm 4.63)	0.98(0.98,0.99)	0.95(0.94, 0.96)
Section C						
- Range (Min, Max)	3.7(3.7, 7.4)	3.7(3.7, 7.4)	3.7(3.7, 7.4)	3.7(3.7, 7.4)		
- Mean(\pm SD)	5.30(\pm 1.84)	5.32(\pm 1.84)	5.30(\pm 1.84)	5.30(\pm 1.84)	0.99(0.99,0.99)	0.97(0.96, 0.98)
Section D						
- Range (Min, Max)	18(2, 20)	18(2, 20)	18(2, 20)	18(2, 20)		
- Mean(\pm SD)	15.87(\pm 4.30)	16.13(\pm 4.12)	15.63(\pm 4.09)	15.88(\pm 4.17)	0.96(0.95,0.97)	0.90(0.87, 0.92)
Total Score						
- Range (Min, Max)	37.2(59.2, 96.4)	38.2(58.2, 96.4)	37.2(59.2, 96.4)	38.2(58.2, 96.4)		
- Mean(\pm SD)	85.59(\pm 10.16)	85.65(\pm 10.25)	85.44(\pm 10.03)	85.56(\pm 10.13)	0.99(0.99,0.99)	0.97(0.96, 0.98)
Subgroups						
Female (n=20)						
- Range (Min, Max)	33.7(62.7, 96.4)	35.7(60.7, 96.4)	32.7(63.7, 96.4)	35.7(60.7, 96.4)		
- Mean(\pm SD)	84.77(\pm 11.50)	85.62(\pm 10.95)	85.87(\pm 10.69)	85.42(\pm 10.88)	0.99(0.97, 0.99)	0.96(0.92, 0.98)
Male (n=170)						
- Range (Min, Max)	37.2(59.2, 96.4)	38.2(58.2, 96.4)	37.2(59.2, 96.4)	38.2(58.2, 96.4)		
- Mean(\pm SD)	85.68(\pm 10.02)	85.66(\pm 10.20)	85.39(\pm 9.98)	85.58(\pm 10.05)	0.99(0.99, 0.99)	0.97(0.97, 0.98)
Indigenous (n=89)						
- Range (Min, Max)	26.2(70.2, 96.4)	26.2(70.2, 96.4)	26.2(70.2, 96.4)	26.2(70.2, 96.4)		
- Mean(\pm SD)	90.45(\pm 5.46)	90.54(\pm 5.70)	90.21(\pm 5.46)	90.40(\pm 5.52)	0.99(0.99, 0.99)	0.98(0.97, 0.98)
Non-Indigenous(n=101)						
- Range (Min, Max)	37.2(59.2, 96.4)	38.2(58.2, 96.4)	37.2(58.2, 96.4)	38.2(58.2, 96.4)		
- Mean(\pm SD)	80.06(\pm 11.40)	80.11(\pm 11.41)	80.03(\pm 11.24)	80.07(\pm 11.31)	0.97(0.95, 0.98)	0.91(0.87, 0.93)
ID Screening Threshold						
- Possible ID (<85)	61(32.1%)	61(32.1%)	59(31.0%)	-		
- No ID (\geq 85)	129(67.9%)	129(67.9%)	131(69.0%)	-	Kappa (3 raters)	0.95

ICC: Inter-class correlation coefficient

DISCUSSION

To our knowledge this is the first study to estimate the IRR of the HASI screening tool for the identification of possible ID in a sample of prisoners. When the same screening observations were blindly scored by three raters, the HASI exhibited excellent IRR on the scaled sections and total scaled score. As expected, the ICC estimates for individual consistency of agreement were lowest for section D (clock-drawing) as this is the most subjective measure to score. Pre-scoring training should pay particular attention to scoring the clock-drawing to maximise IRR. Furthermore, almost perfect agreement was observed when the dichotomous adult ID threshold was evaluated. Our findings indicate that the HASI has a low risk of inter-rater bias at every level: section scores, total scaled score, and the dichotomous adult ID screening outcome.

Our findings must be understood in the context of several limitations. Reliability was assessed by scoring the HASI after administration of the screening tool. Our raters had extensive practical and research experience in prisoner health; the reliability of HASI scoring between raters with no research background is worthy of exploration. The validity of the HASI was not assessed, and this may be subject to other forms of information bias (e.g., cultural bias). However, reliability can be considered a prerequisite for validity; insufficient reliability has been shown to affect the accuracy of measurements (Shrout and Yager 1989). Thus, our findings provide some indication that the validity of the HASI (Hayes 2000) is likely maintained when scored by different raters. Finally, it was outside the scope of this study to investigate intra-rater and test-retest reliability, however both represent important areas for future research.

Despite these limitations, the current study provides new evidence that the HASI has a low risk of bias from inter-rater scoring and can be reliably scored by non-clinicians and clinicians with basic training in research-team settings. Our findings also suggest that HASI scoring is similarly reliable regardless of the gender of participants. Interestingly, the IRR

was better in Indigenous than in non-Indigenous participants. This difference cannot be accounted for by participant characteristics, to which the raters were blind. Scoring in both subgroups exhibited excellent consistency of agreement, such that this difference is of little consequence in most settings. However, some uncertainty arose as to the construct (cross-cultural) validity of using clocks to assess cognitive deficits in Indigenous prisoners from remote communities where the concept and importance of time differs greatly from Western culture: an important area for future consideration.

Our use of individual-level correlation data to assess within-subject consistency of agreement can be considered a strength of our methodology. This indicates that the HASI can be scored reliably by different raters for each individual assessed. A low risk of bias is important as accurate ID screening is necessary to ascertain and identify need within prison and post-release. The reliable scoring observed across three raters from different backgrounds suggests that these results may be generalised to raters from diverse research and clinical backgrounds, with moderate experience and basic pre-scoring training. Similar results would be expected when the HASI is administered in other marginalised groups.

Our findings are particularly relevant in the context of a major funding, policy and practice shift associated with the Australian National Disability Insurance Scheme (NDIS), which will attract over \$AUD 14.3 billion over the next seven years (National Disability Insurance Agency 2013). The NDIS will fund support for eligible persons with ID on remand, community based orders and following release from prison, including assistance with planning, decision making, scheduling, communication and community living. The targeted clinical identification of people with ID through accurate screening by correctional systems will facilitate the generation of data to inform and support interaction between the NDIS and the justice system in Australia.

REFERENCES

- AIHW (2013) The health of Australia's prisoners 2012. Australian Institute of Health and Welfare, Canberra. <http://www.aihw.gov.au/publication-detail/?id=60129543948>
- Baldry, E., Clarence, M., Dowse, L. & Trollor, J. (2013) Reducing Vulnerability to Harm in Adults With Cognitive Disabilities in the Australian Criminal Justice System. *Journal of Policy and Practice in Intellectual Disabilities*, **10**, 222-229.
DOI:10.1111/jppi.12039
- Baldry, E., Dowse, L. & Clarence, M. (2012) People with intellectual and other cognitive disability in the criminal justice system. In: *Report for NSW Family and Community Services Ageing, Disability and Home Care*.
https://www.adhc.nsw.gov.au/data/assets/file/0003/264054/Intellectual_and_cognitive_disability_in_criminal_justice_system.pdf
- Butler, T. & Milner, L. (2003) The 2001 New South Wales Inmate Health Survey. Corrections Health Service Sydney. <http://www.justicehealth.nsw.gov.au/publications/inmate-health-survey-2001.pdf>
- Hayes, S., Shackell, P., Mottram, P. & Lancaster, R. (2007) The prevalence of intellectual disability in a major UK prison. *British Journal of Learning Disabilities*, **35**, 162-167.
DOI:10.1111/j.1468-3156.2007.00461.x
- Hayes, S. C. (2000) Hayes Ability Screening Index (HASI). Manual, University of Sydney: Department of Behavioural Sciences in Medicine, Sydney.
- Hayes, S. C. (2002) Early Intervention or Early Incarceration? Using a Screening Test for Intellectual Disability in the Criminal Justice System. *Journal of Applied Research in Intellectual Disabilities*, **15**, 120-128. DOI:10.1046/j.1468-3148.2002.00113.x
- Indig, D., Topp, L., Ross, B., Mamoon, M., Border, B., Kumar, S. & McNamara, M. (2010) 2009 NSW Inmate Health Survey: Key Findings Report. Justice Health NSW, Sydney. <http://www.justicehealth.nsw.gov.au/publications/2009-ihs-report.pdf>

National Disability Insurance Agency (2013) National Disability Insurance Scheme: Federal Budget. www.ndis.gov.au

Rousson, V., Gasser, T. & Seifert, B. (2002) Assessing intrarater, interrater and test–retest reliability of continuous measurements. *Statistics in Medicine*, **21**, 3431-3446.

DOI:10.1002/sim.1253

Shrout, P. E. & Fleiss, J. L. (1979) Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, **86**, 420. DOI:10.1037/0033-2909.86.2.420

Shrout, P. E. & Yager, T. J. (1989) Reliability and validity of screening scales: Effect of reducing scale length. *Journal of Clinical Epidemiology*, **42**, 69-78.

DOI:10.1016/0895-4356(89)90027-9

Simpson, J., Green, J. & Martin, M. (2001) The framework report: Appropriate community services in NSW for offenders with intellectual disabilities and those at risk of offending. New South Wales Council for Intellectual Disability.

www.idrs.org.au/pdf/historic/fw_main.pdf

StataCorp (2013) Stata Statistical Software: Release 13. StataCorp LP, College Station, TX.

<http://www.stata.com/>