

Validation and Comparability of Smartphone-based Digit Symbol Substitution Task with Written Version

Allen L¹, Baker E¹, Thorp E¹, Evans M², Thorpe D², Granger K¹, Cormack F¹, Barnett J^{1,3}, Cashdollar N⁴

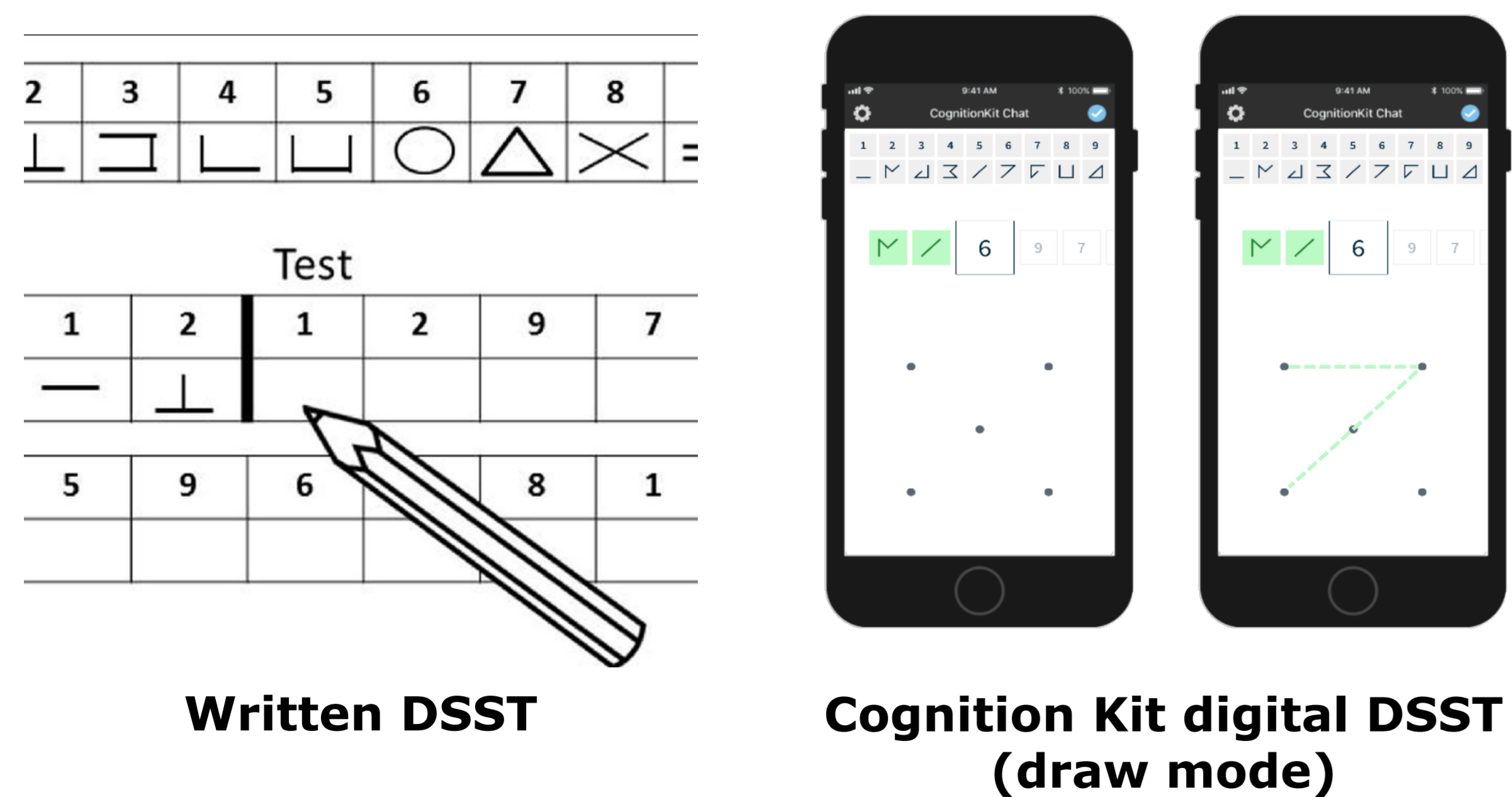
¹Cambridge Cognition, Cambridge, UK, ²Adams Clinical, Watertown, MA, USA, ³University of Cambridge, Cambridge, UK, ⁴Cambridge Cognition, Cambridge, MA, USA

Background

- Digit-symbol substitution tasks (DSST) are widely utilised cognitive tests which provide objective measures of global cognitive function, motor impairment and processing speed. DSST is traditionally administered in written format (paper-pencil), with digital versions emerging but largely not implemented. Such digital tasks require validation against traditional forms and in clinical populations.
- In collaboration with Adams Clinical, participants with major depressive disorder (MDD) underwent both written (WAIS III¹) and digital versions of the DSST offering a unique opportunity to inform comparability and validity of smart phone delivered digital DSST; Cognition Kit DSST.

Methods

- Participants aged 18-85 years meeting DSM-V criteria for MDD, with a Montgomery-Asberg Depression Rating Scale score of ≥ 20 were recruited into an open-label treatment study and underwent both smartphone-based and written DSST assessments. Only right-handed participants were considered for comparability analyses, since few left handed (n=17) and ambidextrous (n=5) participants were recruited.
- Agreement between task versions at baseline and visit 2 were evaluated using Pearson and intra-class correlation coefficients (ICC), as well as Lin's concordance coefficients (CCC), Bland-Altman plots and Passing-Bablok regression plots. Test-retest reliability was assessed with ICCs in a sub-group of participants who completed assessments at baseline and 28 days later. Associations with demographic data such as age and sex were also explored.



Results

N participants	Age \pm SD	Sex (F:M)	MADRS \pm SD
89	32.94 \pm 11.42	59:30	33.16 \pm 4.57

- Moderate to good agreement was achieved between written and digital DSST at baseline (ICC=0.69, $r=0.70$, $p=1.389e-14$ - Fig 1L), and visit 2 (ICC=0.60, $r=0.61$, $p=0.0003$ Fig 2L). Adjusting for order (version completed first) using linear regression, improved agreement to ICC=0.78 at baseline, and ICC=0.68 at visit 2.
- Bland-Altman and Passing-Bablok regression plots revealed an upward bias in written DSST (19.88 extra correct items - Fig 1 & 3).
- Concordance between digital and written scores improved after adjustment of digital scores (by adding average score difference of 19.88 to digital scores; CCC=0.25 to 0.70; baseline - Fig 1).
- Comparable correlations were found between age and digital ($r=-0.47$), and written ($r=-0.41$) scores at baseline.
- A small positive correlation was found for digital between-visit score difference vs written between-visit score difference ($r=0.25$, $p=0.19$; Fig 2R).
- Test-retest reliability was comparable between written and digital measurements, albeit slightly stronger for digital (ICC=0.87 vs ICC=0.84; Fig 4).

Conclusions

- This work supports the validity of Smartphone-based DSST for assessments in patient populations such as MDD. Higher written scores likely reflect the smaller movements required when manually writing symbols, compared with drawing via the Smartphone screen. Future work will focus participant usability of digital DSST and sensitivity to clinical features and change over time.

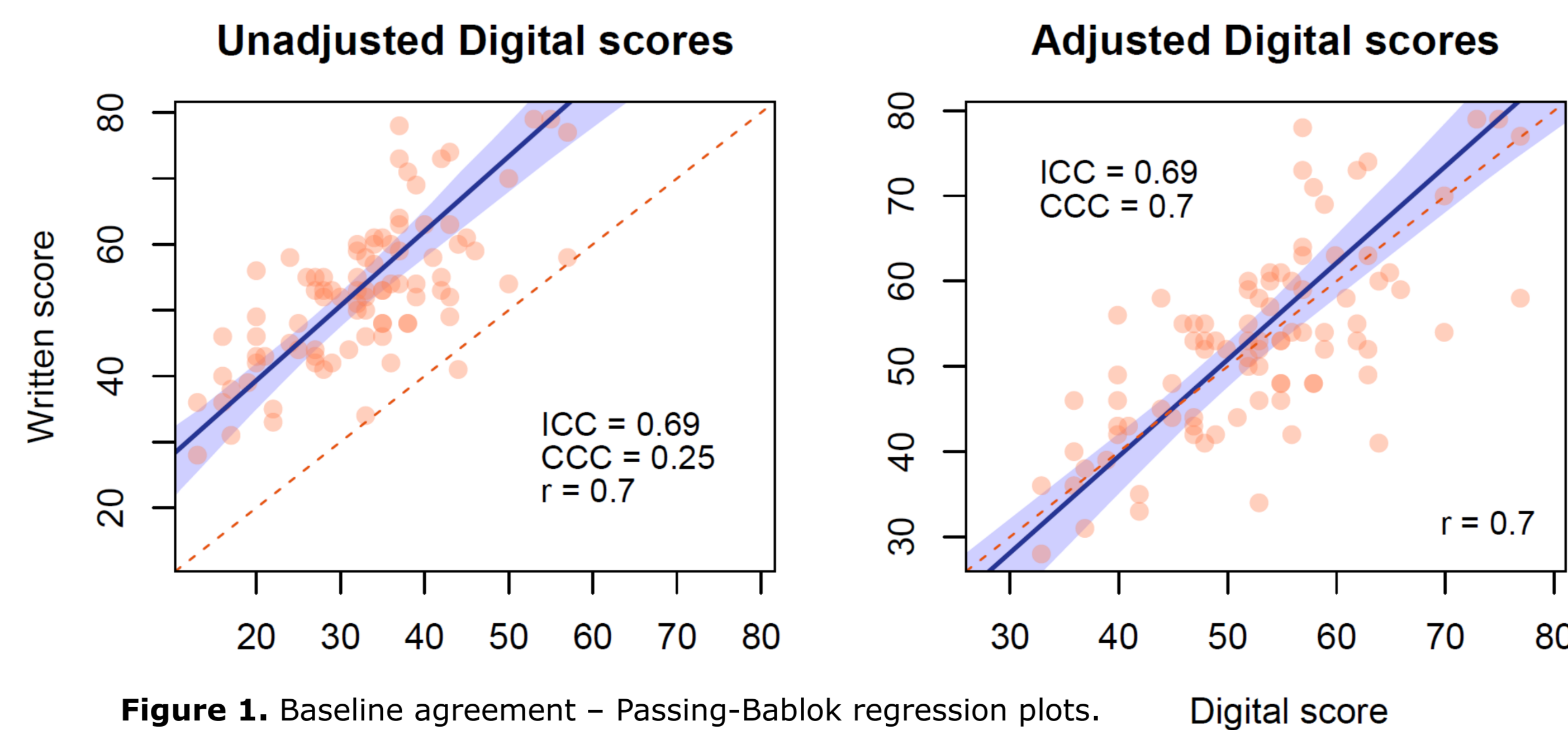


Figure 1. Baseline agreement - Passing-Bablok regression plots. Left: written vs *unadjusted* digital scores. Right: written vs *adjusted* digital scores. (n=89)

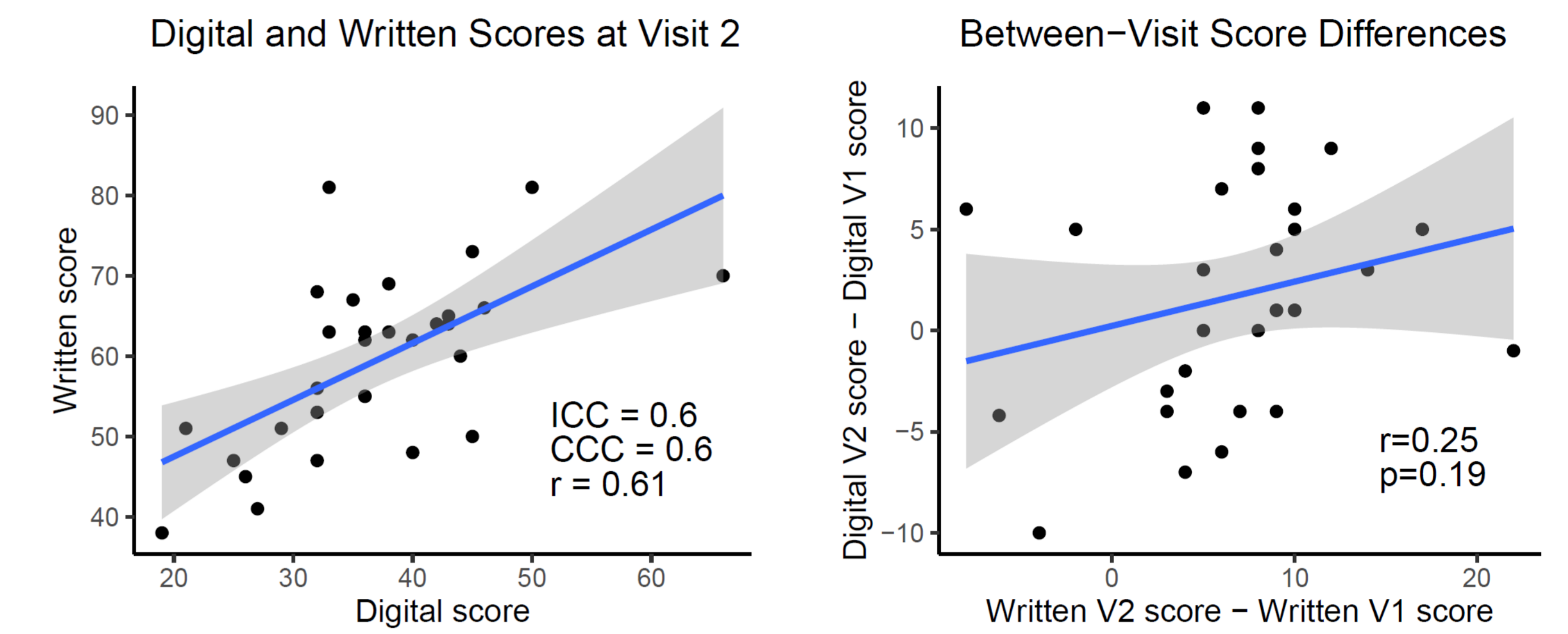


Figure 2. Left: Digital and Written scores at baseline; Right: between-visit score difference agreement

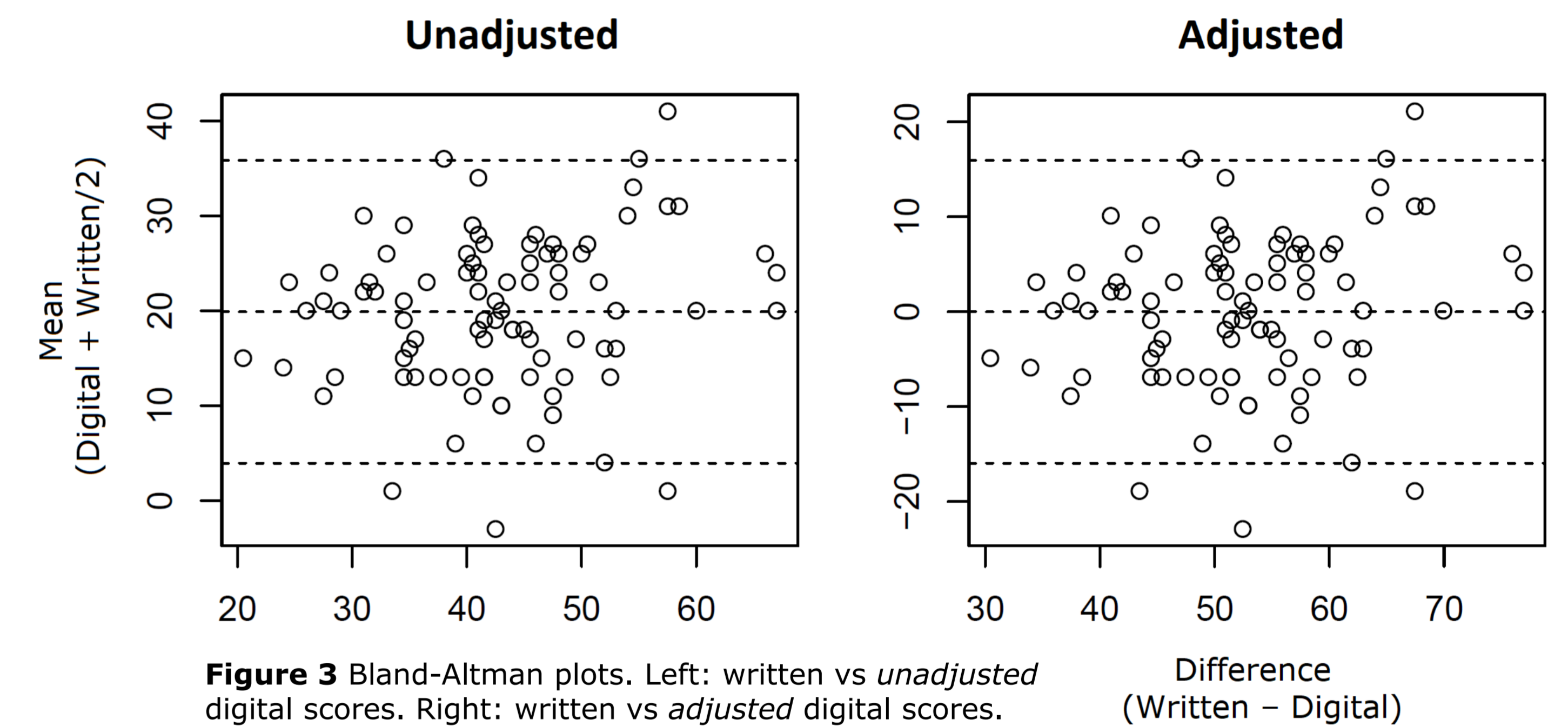


Figure 3 Bland-Altman plots. Left: written vs *unadjusted* digital scores. Right: written vs *adjusted* digital scores. (baseline; n=89)

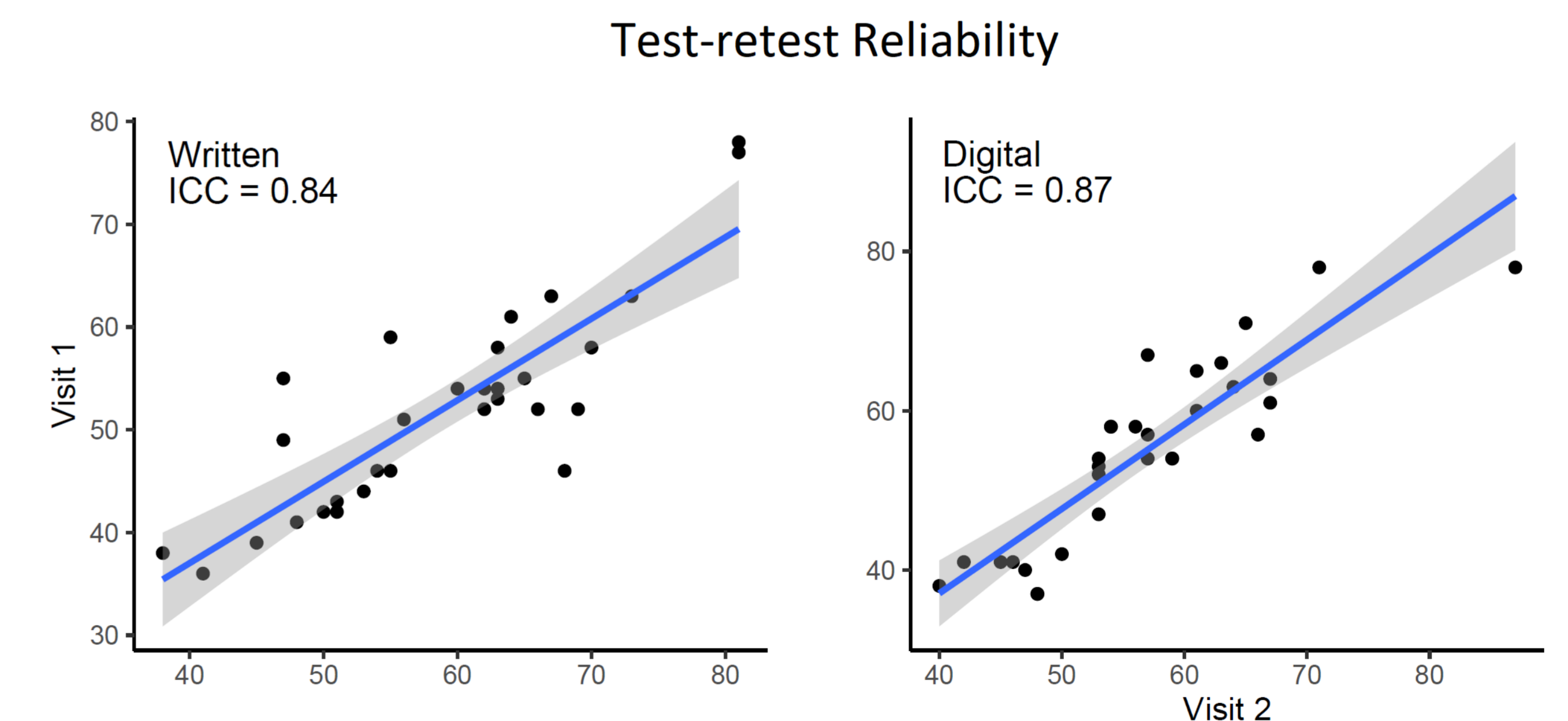


Figure 4 Test-retest scatter plots. Left: written. Right: digital.