

Cultural Influences on Speech-Based Cognitive Assessment: Insights from an Egyptian Cohort

Simona Schäfer¹, Alexandra König¹, Irene Meier², Vaibhav Narayan², Sara A. Moustafa³, Salma Mowafi³, Mohamed Salama³

¹ki:elements GmbH, Saarbrücken, Germany; ² Davos Alzheimer's Collaborative; ³ Institute of Global Health and Human Ecology, American University in Cairo, Cairo, Egypt

Introduction

Dementia is a global health crisis, yet clinical trials often lack diversity due to insufficient neuropsychological tests adapted for cultural and educational variations. Beyond education, individual differences in test-taking behavior (e.g., preferences for accuracy vs. speed, long-term vs. short-term orientation) can significantly influence results. This study investigates how education and specific test-taking behaviors affect performance on the Speech Biomarker for Cognition (SB-C) in an Egyptian cohort, with the goal of ensuring equitable diagnostic utility across diverse populations.

Methods

285 Arabic-speaking participants in Egypt were analysed (see Table 1). Cognitive status (Healthy Control [HC] vs. Dementia) was established using the Harmonized Cognitive Assessment Protocol (HCAP). Participants completed speech tasks, including the Rey Auditory Verbal Learning Test (RAVLT) and Semantic Verbal Fluency (SVF; administered twice), to generate a composite Speech Biomarker for Cognition (SB-C). SB-C score and subscores were derived using ki:elements' proprietary speech analysis pipeline. Analyses compared SB-C scores across three educational levels (illiterate, 1–7 years, 8+ years). Performance instability categories were built. The SVF was performed twice. We built two groups based on the performance change between the first and the second trial. Due to practice effects, an increase would have been expected. Participants who showed a similar or improved performance were labeled as "stable/improving" and participants who declined in their performance were labeled "unstable/declining". Kruskal-Wallis tests examined group differences (HC vs. Dementia) across education and performance stability groups.

Table 1. Participant demographics and group characteristics (N=285)

	HC (N = 158)	Dementia (N = 127)	p-value
Sex (F / M)	62 / 96	97 / 30	< .001***
Age, mean (SD)	64.3 (8.0)	65.2 (7.2)	n.s.
Education, mean (SD)	4.8 (6.5)	2.3 (4.8)	< .05*

Figure 1. SB-C Cognition Score by Education Group (HC vs. Dementia, Kruskal-Wallis, FDR-corrected, N=285)

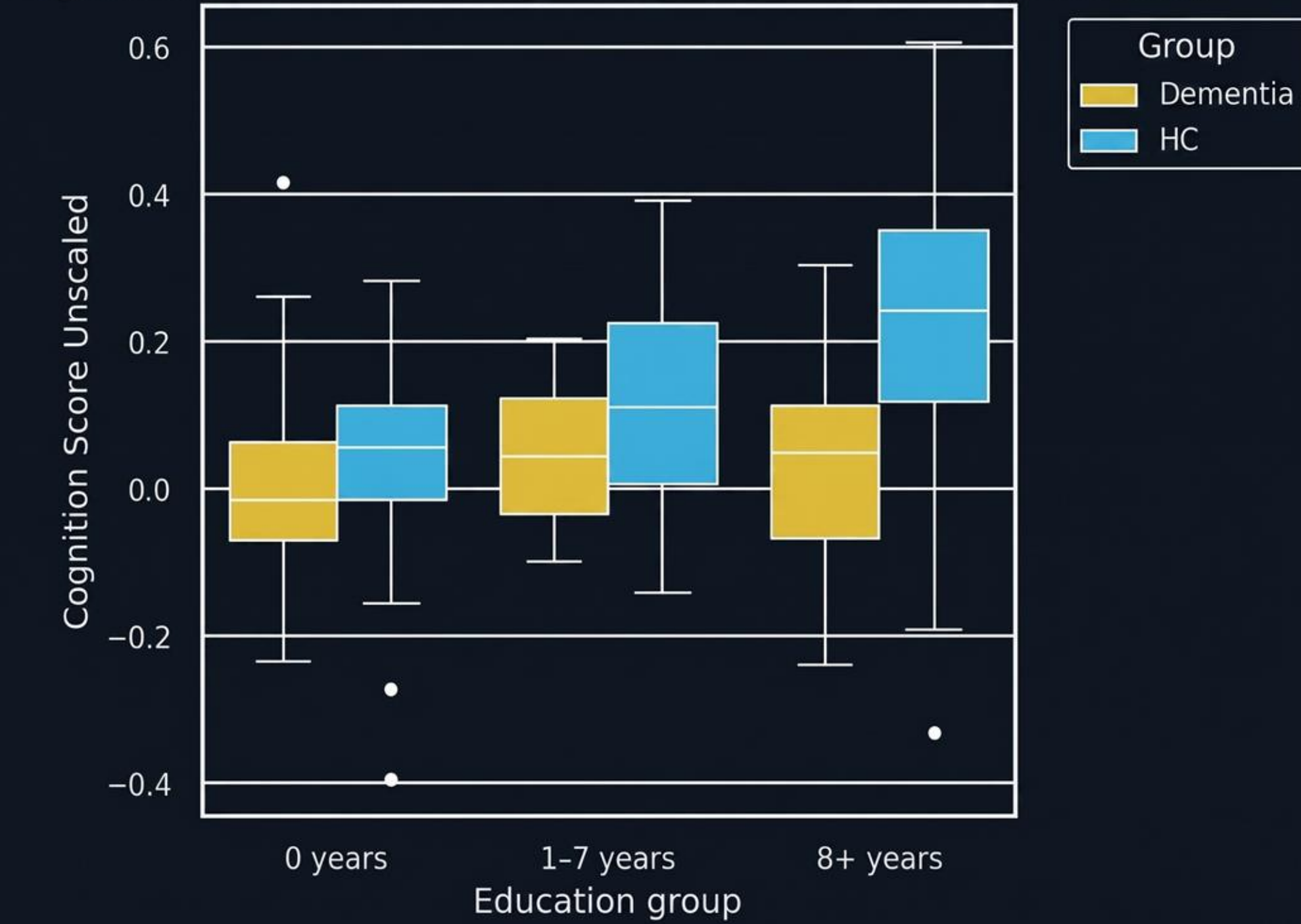


Figure 2. SB-C Subdomain Scores by Diagnostic Group (HC vs. Dementia, Kruskal-Wallis H=38.27, p<.001, N=285)

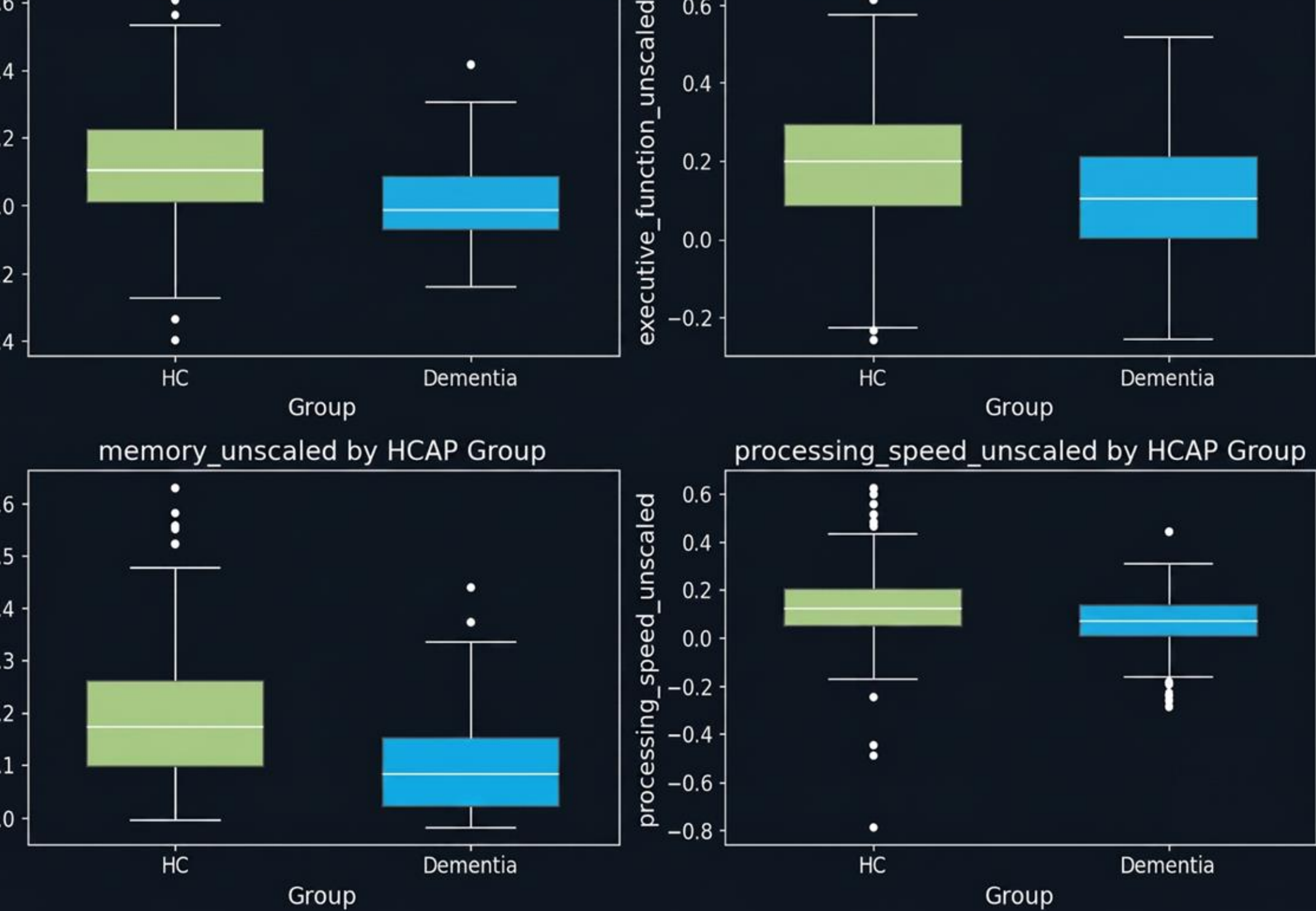
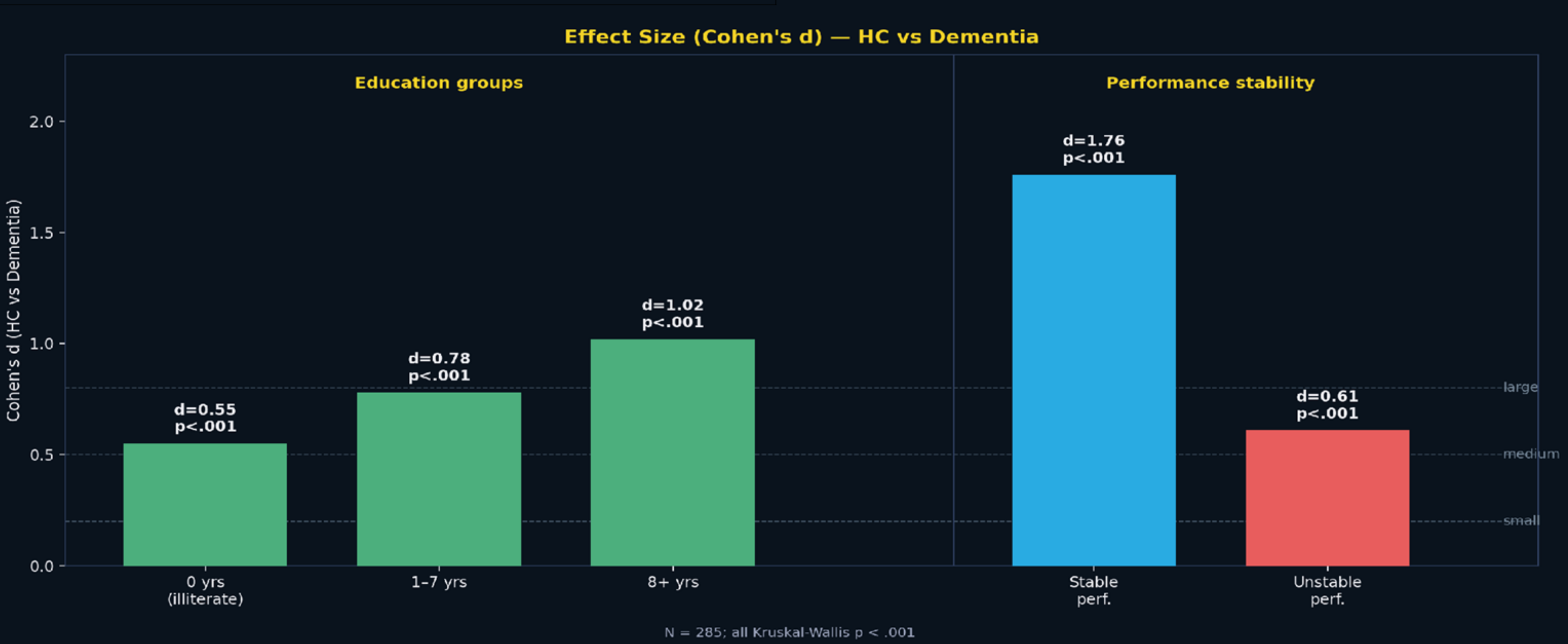


Figure 3. Effect Size (Cohen's d) by Education Group and SVF Performance Stability (HC vs. Dementia, N=285)



Results

Educational level significantly influenced assessment sensitivity. In the illiterate group, the SB-C showed a moderate difference between HC and Dementia participants (H = 13.94, $p < .001$, $d = 0.55$). The separation between clinical groups widened with increasing education (8+ years: H = 14.83, $p < .001$, $d = 1.02$; see Fig. 1). Regarding test-taking behavior, some participants performed worse on SVF repetition — contrary to expected learning effects. Group differences were significantly larger in the stable/improved performance group (H = 23.3, $p < .001$, $d = 1.76$) compared to the unstable/declined performance group (H = 20.86, $p < .001$, $d = 0.61$; see Fig. 2).

Conclusion

Education and test-taking behavior directly impact the diagnostic utility of speech-based assessments. Lower education levels reduced clinical group separation, likely due to floor effects. Performance instability significantly diminished diagnostic accuracy. To address global healthcare disparities, speech-based algorithms must be calibrated for sociodemographic and behavioral variables to ensure equitable precision across diverse populations.

Contact

alexandra.koenig@ki-elements.de

