Assessing the utility of electronic measures as a proxy for cognitive ability

Tess Levy, Bari Britvan, Hannah Grosman, Ivy Giserman-Kiss, Kristin Meyering, Jordan Weissman, Danielle Halpern, Jennifer Foss-Feig, Alexander Kolevzon, M. Pilar Trelles, Stephan Sanders, Elise Robinson, Joseph Buxbaum, Somer Bishop, Paige M. Siper

Abstract

Large-scale genomic studies have identified over 100 genes associated with autism spectrum disorder (ASD); however, important phenotypic variables are captured inconsistently. In many cases, the resources required for comprehensive characterization hinder the feasibility of collecting critical information, such as intellectual ability. Thus, electronic collection of important phenotypes would greatly facilitate large-scale data collection efforts. This study assessed the utility of two electronic assessments as a proxy of cognitive ability relative to clinician-administered cognitive assessments. Ninety-two participants completed the study, including individuals with ASD (probands, n = 19), parents of probands (n = 46), and siblings without ASD (n = 27). Participants were administered the electronic Peabody Picture Vocabulary Test, Fourth Edition (e-PPVT-4), an electronic visual reasoning (VR) test, and a clinician-administered Wechsler Abbreviated Scales of Intelligence, Second Edition (WASI-II). Proband also completed a full, in-person, cognitive assessment and Vineland Adaptive Behavior Scales, 2nd Edition. Correlations between scores on electronic and clinician-administered measures were examined. Classification accuracy of individual scores based on 95% confidence intervals and score range (below average, average, above average)
were also assessed. Moderate to strong correlations were identified between both electronic measures and the clinician-administered WASI-II ($\rho = 0.606–0.712$). Mean difference between standard scores ranged from 10.7 to 14.8 for the cohort. Classification accuracy based on WASI-II 95% confidence interval was consistently low (27.5%–47.3%). Classification accuracy by score range (below average, average, above average) was variable, ranging from 33% to 86% for probands. All participants unable to complete the electronic assessments met DSM-5 criteria for intellectual disability. e-PPVT-4 and VR scores were strongly correlated with scores on the WASI-II full-scale IQ ($\rho = 0.630, 0.712$), indicating utility of these measures at the group level in large-scale genomic studies. However, the poor precision of measurement across both measures suggests that the e-PPVT-4 and VR are not useful alternatives to in-person testing for the purpose of clinical assessment of an individual’s IQ score.

**Lay Summary**

Large-scale studies designed to identify genes associated with autism have been successful in identifying over 100 genes. However, important clinical information about participants with autism and their family members is often missed—including cognitive functioning. Cognitive testing requires in-person administration by a trained clinician and therefore can be burdensome and often reduces feasibility of diverse samples. Here, we assessed whether electronic assessments could take the place of in-person cognitive testing. We found that at the group level, for large-scale studies, electronic measures added valuable information; however, they were not accurate enough to be used on an individual level (i.e., to offer feedback about an individual’s predicted IQ score).

**KEYWORDS**
cognitive testing, electronic, gene discovery, phenotyping, phenotyping autism spectrum disorder

**INTRODUCTION**

Large-scale genomic studies have been paramount to discovering genes associated with autism spectrum disorder (ASD). The most recent ASD gene discovery study included ~12,000 cases and identified over 100 genes associated with autism (Satterstrom et al., 2020). However, there remains a gap in the ability to systematically obtain critical phenotypic information beyond a categorical diagnosis. The clinical characterization burden, both in time and resources, has hindered the analytic potential of existing genetic databases in which key variables, such as cognition, is inconsistently captured. Additionally, the travel burden for participation in these studies limits sample diversity.

Cognitive ability is one of the strongest indicators of ASD symptom severity and prognosis (Ben-Itzchak et al., 2014; Flanagan et al., 2012; Harris & Handleman, 2000), and intellectual disability is strongly associated with rare de novo variants (Brkić et al., 2020; Iossifov et al., 2014; O’Roak et al., 2012). Identifying ways to increase feasibility of intellectual quotient (IQ) measurement is highly relevant to the field. While clinician-administered cognitive assessments are considered the gold-standard, several “next-best” options exist that can provide a proxy of cognitive ability and require fewer resources. Abbreviated cognitive assessments, including the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011) and the Kaufman Brief Intelligence Test (Kaufman, 1990), are fair alternatives for measuring general intellectual functioning (Canivez et al., 2009; Hays et al., 2002). However, administration still requires an in-person visit and trained examiner. An electronic delivery platform would increase sample diversity. For example, Simons Foundation Powering Autism Research (SPARK; Feliciano et al., 2018), an autism genetics study, allows families to participate remotely by mailing in saliva samples, however, estimates of cognitive ability are not prospectively captured. Identifying a proxy for cognitive variability via remote assessments would be a notable advance. Additionally, genetic studies often use “trio” analysis by including both parents of the proband, or siblings when both parents are not available. Data on family member controls’ cognitive ability is another rarely available, relevant datapoint in analyzing and interpreting genomic data. We aimed to measure the viability of in-person electronic assessments with the goal of eventually transitioning to remote administration. We assessed the accuracy of two electronic tests, the electronic Peabody Picture Vocabulary Test, Fourth Edition (Dunn & Dunn, 2007) to assess verbal ability and an electronic Visual Reasoning test (Germine et al., 2012) to assess nonverbal ability, in predicting scores on a clinician-administered cognitive assessment in individuals with ASD and their immediate family members.
METHODS

Participants

Ninety-seven individuals including probands, parents, and typically developing siblings enrolled in this study. Only participants who completed the WASI-II and at least one of the electronic measures were included in analyses. Five probands were excluded because they did not obtain basal scores on WASI-II subtests. These five probands were unable to complete the e-PPVT-4, and four of five were unable to complete the VR. The final sample included 92 participants: 19 probands 6–24 years old (10.68 ± 4.9), 46 parents (46.5 ± 7.0), and 27 siblings (11.56 ± 4.8; Table 1). This study was approved by the Mount Sinai Program for the Protection of Human Subjects. Informed consent was obtained for all participants, and assent was obtained when appropriate.

Procedure

Participants completed (a) a clinician-administered WASI-II (n = 92; Wechsler, 2011), (b) an electronic version of the PPVT-4 (e-PPVT-4; n = 92; Dunn & Dunn, 2007) and (c) an electronic Visual Reasoning test (VR; n = 91) developed by TestMyBrain.org (Chaytor et al., 2021; Germine et al., 2012; Singh et al., 2021). Electronic assessments were completed in the clinic (2017–2019) on a 15-in. laptop with an external mouse, which was used to select response on both tests. Testing was supervised by a research coordinator who was responsible for setting up the assessments. The coordinator did not assist or prompt in any capacity. The two electronic assessments were completed in the same block of time, with the order alternated for each participant. The WASI-II full-scale IQ was calculated based on all four subtests. WASI-II verbal comprehension and perceptual reasoning (nonverbal) indices were also calculated. The e-PPVT-4 is a standardized assessment of receptive vocabulary where individuals were verbally instructed to point to pictures of specific objects, actions, and concepts, among four choices. The e-PPVT-4 yields an overall standard score. The VR test, which is made up of matrix reasoning tasks, was chosen for the electronic measure of nonverbal ability; matrix reasoning tasks are among the best measures of fluid and general intelligence (Carroll, 1997). The VR test was designed to be a measure of cognitive functioning, and specifically of attention, perception, cognitive control, and working memory (Passell et al., 2019). It has similar reliability to the WASI-II Matrix Reasoning test (Cronbach’s alpha = 0.77; Passell et al., 2019). Additionally, the VR test correlates with the scholastic aptitude test (SAT) Math (rho = 0.41, n = 1345, 95% CIs [0.37, 0.45]); and to a lesser degree with Verbal (rho = 0.22, n = 1358, 95% CIs [0.17, 0.27]) and Vocabulary (rho = 0.31, n = 10,000, 95% CIs [0.29, 0.33]; Passell et al., 2019). The VR test asked participants to solve visual puzzles by selecting an answer among five options to complete a visual pattern. Z-scores were calculated via norms provided by TestMyBrain, and then used to calculate standard scores.

In addition to the WASI-II, probands were administered a full clinician-administered cognitive assessment based upon age and functioning level; participants received the Wechsler Intelligence Scales for Children, Fifth Edition (n = 13; Wechsler, 2014),

TABLE 1 Cohort demographics and descriptive statistics

|                         | All participants | Probands | Parents | Siblings |
|-------------------------|------------------|----------|---------|----------|
| Sample size             | 92               | 19       | 46      | 27       |
| Age, years              | 28.6 (18.7)      | 10.7 (4.9) | 46.5 (7.0) | 11.6 (4.8) |
| % female                | 54.3%            | 39.7%    | 50%     | 74%      |
| WASI-II full-scale IQ   | 108.79 (17.3)    | 90.21 (23.6) | 112.37 (11.1) | 115.78 (11.1) |
| WASI-II verbal comprehension index | 109.34 (18.9) | 87.74 (23.6) | 114.04 (13.1) | 116.50 (11.8) |
| WASI-II perceptual reasoning index | 105.68 (15.8) | 93.63 (22.0) | 107.30 (10.9) | 111.40 (13.7) |
| e-PPVT-4 standard score | 102.02 (15.9)    | 87.58 (23.4) | 103.78 (9.9) | 109.19 (11.0) |
| VR standard score       | 100.16 (14.2)    | 87.44 (17.0) | 100.6 (11.6) | 101.2 (14.9) |
| SRS-2 Total T-Score     | 53.34 (14.9)     | 76.47 (11.7) | 48.80 (8.8) | 44.63 (5.9) |
| Vineland-II adaptive behavior composite | n/a            | 72.37 (11.4) | n/a      | n/a      |
| ADOS-2 comparison       | n/a              | 6.78 (1.8) | n/a     | n/a      |
| Full cognitive assessment full-scale IQ | n/a          | 93.56 (24.7) | n/a    | n/a      |
| Full cognitive assessment verbal IQ | n/a        | 94.83 (24.6) | n/a   | n/a      |
| Full cognitive assessment nonverbal IQ | n/a        | 94.71 (29.1) | n/a   | n/a      |

Note: Scores shown as mean (standard deviation).
Abbreviations: e-PPVT-4: electronic Peabody Picture Vocabulary Test, Fourth Edition; IQ: Intellectual quotient; Probands: individuals with autism spectrum disorder; VR: Visual Reasoning test; WASI-II: Wechsler Abbreviated Scales of Intelligence, Second Edition.
Wechsler Adult Intelligence Scales, Fourth Edition \((n = 2;\) Drozdick et al., 2018), Stanford-Binet Intelligence Scales, Fifth Edition \((n = 3;\) Roid & Pomplun, 2012), or the Mullen Scales of Early Learning \((n = 1).\) The Vineland Adaptive Behavior Scales, Second Edition (Sparrow, 2011; Sparrow et al., 2005) was administered to all probands to assess ID, as deficits in cognitive and adaptive functioning are required for a diagnosis of ID.

Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (American Psychiatric Association, 2013) ASD diagnoses were confirmed based on an evaluation with a board-certified psychiatrist or licensed psychologist, incorporating information from the Autism Diagnostic Observation Schedule, Second Edition \((n = 18;\) Lord et al., 2012) and the Autism Diagnostic Interview-Revised \((n = 8;\) Lord et al., 1994) or the Social Communication Questionnaire \((n = 11;\) Rutter et al., 2003). All 19 probands had a confirmed DSM-5 ASD diagnosis. Parents and siblings were screened for ASD using the Social Responsiveness Scale-2nd Edition (SRS-2); and none were excluded.

**Analysis**

Spearman’s rank-order correlation coefficients \((\rho)\) were used to examine the relationship between WASI-II full-scale IQ and both the e-PPVT-4 and VR. Correlations between the e-PPVT-4 and WASI-II verbal comprehension index and between the VR and WASI-II perceptual reasoning index were assessed. Mean difference (absolute value) between standard scores of an individual’s e-PPVT-4 or VR score with their WASI-II score was calculated. The proportion of cases in which e-PPVT-4 or VR fell within an individual’s WASI-II full-scale IQ 95% confidence interval was evaluated. Additionally, to assess if scores on the e-PPVT-4 or VR predicted classification ranges on the WASI-II full-scale IQ, groups were divided based on scores under 85 (below average), scores 85–115 (average), and scores over 115 (above average). Analyses assessed the electronic measures with full-scale IQ to examine a potential relationship with overall IQ, as well as with the respective verbal/nonverbal index. Analyses were completed on the whole cohort and within family member type.

**RESULTS**

The WASI-II and e-PPVT-4 were completed by 92 participants, and the VR by 91 participants (Figure 1). Mean scores (Table 1) were significantly lower in probands relative to both parents and siblings on the WASI-II \((p = 0.002, p < 0.001),\) e-PPVT-4 \((p = 0.003, p < 0.001),\) and VR \((p = 0.002, p = 0.013).\) There was no significant difference between parent and sibling scores.

**Correlations between WASI-II and electronic assessments**

WASI-II full-scale IQ was strongly correlated with both e-PPVT-4 \((\rho = 0.712, p < 0.001)\) and VR \((\rho = 0.630, p < 0.001;\) Table 2, Figure 2a). The e-PPVT-4 was strongly correlated with the WASI-II verbal comprehension index \((\rho = 0.685, p < 0.001)\) and the VR was strongly correlated with the WASI-II perceptual reasoning index \((\rho = 0.608, p < 0.001).\)

**Mean difference between WASI-II and electronic measures**

A comparison of the discrepancy in standard scores between the e-PPVT-4 and VR with the WASI-II were calculated for each participant. The mean difference between e-PPVT-4 and full-scale IQ was 10.7, and the mean difference for the VR and full-scale IQ was 13.4, where both electronic measures tended to yield lower scores than in-person testing (Table 3, Figure 2b).

**95% confidence interval accuracy**

The proportion of cases that e-PPVT-4 and VR scores accurately classified based on WASI-II 95% confidence intervals are displayed in Table 3. Participant’s e-PPVT-4 and VR score were in the WASI-II full-scale IQ 95% confidence interval 30.4% and 27.5% of cases, respectively.
The e-PPVT-4 correctly classified WASI-II full-scale IQ score ranges in 71% of below average scores, 57% of average scores, and 69% of above average scores (Table 3, Figure 2c). The VR correctly classified WASI-II scores in 33% of below average, 59% of average, and 86% of above average cases. Classification accuracy by participant group is displayed in Table 3.

**DISCUSSION**

This study explored the potential of measuring cognition using electronic assessments. Two computer-based measures, the e-PPVT-4, a measure of receptive vocabulary, and the VR test, a measure of attention, perception, cognitive control, and working memory, were administered to individuals with autism and their immediate family members and compared to clinician-administered IQ
tests. The e-PPVT-4 and VR showed moderate to strong correlations with full-scale IQ on an abbreviated measure of intellectual functioning (WASI-II), and with verbal and perceptual reasoning indices, respectively, suggesting measurement of similar constructs. This highlights the utility of these measures in large scale genomic studies. However, scores on electronic measures were nearly one standard deviation below WASI-II scores and only fell within the WASI-II 95% confidence interval less than a third of the time, indicating poor precision of measurement. The e-PPVT-4 better predicted below average scores (71%) compared to the VR test (33%) and both tests similarly predicted IQs within the average range (57%–59%). Results indicate the electronic measures should not be used clinically as a proxy for IQ.

With the goal of increasing available data for genomic studies, inclusion of information from profoundly affected individuals is critical. However, 20% of probands (5/24) were unable to complete the electronic measures due to functional capacity; all these participants received a diagnosis of intellectual disability. If other factors are ruled out (e.g., attention, motor), incomplete results may be a useful indicator of low cognitive ability. Correlations from parents and siblings were weaker than those for probands, likely in part due to the lower variability of scores within groups.

Overall, correlations indicate these electronic assessments adequately relate to functioning on the WASI-II at the group level; however, results lack the precision needed to estimate scores at the individual level. This is consistent with prior literature illustrating poor predictive ability between various IQ measures, despite significant correlations (Bell et al., 2001). Overall, there was not a clear difference in performance between the e-PPVT-4 and VR test; however, the e-PPVT-4 tended to better capture the more severely affected individuals, which is an important consideration. Future studies with larger and more variable samples can further evaluate these electronic assessments.

**LIMITATIONS**

Our sample size was modest, and therefore does not capture the extent of the heterogeneity in real-world ASD populations. Specifically, while a larger proportion (38%) of probands in our originally recruited cohort had ID, many were unable (five of nine individuals with ID) to complete electronic assessments, leaving the resulting cohort with underrepresentation of ID (21% ID compared to 33% expected [Baio et al., 2018; Shaw et al., 2020]). Additionally, electronic measures were completed in the clinic under controlled conditions and results may not generalize to remote administration.

**CONCLUSION**

Electronic assessments show strong correlations with IQ, however, poor precision in estimating absolute IQ scores.
Nevertheless, electronic assessments may provide a useful means to fill gaps within existing autism databases and enhance future large-scale genomic research in the field. In contrast, our results indicate these measures should not be used as a proxy for clinical assessment of individual IQ scores.

ACKNOWLEDGMENTS
The authors would like to thank all of the families who generously participated in this study.

CONFLICT OF INTEREST
Alexander Kolevzon receives research supports from AMO Pharma and consults to Acadia, Alkermes, Neuren, and GW Pharma. He serves on Scientific Advisory Boards for Ovid Therapeutics, Jaguar Therapeutics, and Ritrova Therapeutics. Joseph D. Buxbaum consults to BridgeBio, holds a patent for IGF-1 in Phelan-McDermid syndrome, holds an honorary professorship from Aarhus University Denmark, receives research support from Takeda and Oryzon and is a journal editor for Springer Nature.

AUTHOR CONTRIBUTIONS
Sommer Bishop, Stephan J. Sanders, Elise B. Robinson, Joseph D. Buxbaum, and Paige M. Siper were responsible for the conception and design of the study. Paige M. Siper, Ivy Giserman-Kiss, Danielle Halpern, Jessica Zweifach, Maria del Pilar Trelles, Hannah Grosman, Kristin Meyer, Jordana Weissman, Alexander Kolevzon, and Jennifer H. Foss-Feig acquired the data. Tess Levy, Bari Britvan, Paige M. Siper, Sommer Bishop, Hannah Grosman, Ivy Giserman-Kiss, and Jennifer H. Foss-Feig participated in the analysis and/or interpretation of data. Tess Levy, Bari Britvan, and Paige M. Siper drafted the manuscript. Sommer Bishop, Stephan J. Sanders, Elise B. Robinson, Jennifer H. Foss-Feig, Alexander Kolevzon, Joseph D. Buxbaum, Paige M. Siper, and Tess Levy were responsible for revising the manuscript for critically important intellectual content. All authors have approved the version of the manuscript to be published.

ETHICS STATEMENT
All participants or their legal guardians provided informed consent prior to study participation. Assent from minors was collected when appropriate. The study was approved by the Mount Sinai Program for the Protection of Human Subjects (Study ID: 16-0494).

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID
Tess Levy https://orcid.org/0000-0003-2207-0440
Bari Britvan https://orcid.org/0000-0002-5158-3889
Hannah Grosman https://orcid.org/0000-0002-0965-5553
Ivy Giserman-Kiss https://orcid.org/0000-0002-1687-7638
Kristin Meyer https://orcid.org/0000-0001-7568-1363
Jordana Weissman https://orcid.org/0000-0003-2798-4858
Danielle Halpern https://orcid.org/0000-0001-6745-1197
Jessica Zweifach https://orcid.org/0000-0001-5392-8839
M. Pilar Trelles https://orcid.org/0000-0002-7442-3677
Jennifer H. Foss-Feig https://orcid.org/0000-0002-1491-8248
Alexander Kolevzon https://orcid.org/0000-0001-8129-2671
Stephan J. Sanders https://orcid.org/0000-0001-9112-5148
Elise B. Robinson https://orcid.org/0000-0002-6527-901X
Joseph D. Buxbaum https://orcid.org/0000-0001-8898-8313
Sommer Bishop https://orcid.org/0000-0002-2592-4715
Paige M. Siper https://orcid.org/0000-0002-9659-8232

REFERENCES
American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Publication.
Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., Kurzius-Spencer, M., Zahorody, W., Robinson, Rosenberg, C., White, T., Durkin, M. S., Imm, P., Nikolau, L., Yeargin-Allsopp, M., Lee, L. C., Harrington, R., Lopez, M., Fitzgerald, R. T., Hewitt, A., … Dowling, N. F. (2018). Prevalence of autism Spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2014. Morbidity and Mortality Weekly Report, 67(6), 1–23. https://doi.org/10.15585/mmwr.s6706a1
Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the Peabody picture vocabulary test-third edition and Wechsler adult intelligence scale-third edition with university students. Journal of Clinical Psychology, 57(3), 417–422. https://doi.org/10.1002/jclp.1024
Ben-Izhak, E., Watson, L. R., & Zachar, D. A. (2014). Cognitive ability is associated with different outcome trajectories in autism spectrum disorders. Journal of Autism and Developmental Disorders, 44(9), 2221–2229. https://doi.org/10.1007/s10803-014-2091-0
Brikić, D., Ng-Cordell, E., O’Brien, S., Scerif, G., Astle, D., & Baker, K. (2020). Gene functional networks and autism spectrum characteristics in young people with intellectual disability: A dimensional phenotyping study. Molecular Autism, 11(1), 98. https://doi.org/10.1186/s13229-020-00403-9
Canivez, G. L., Konold, T. R., Collins, J. M., & Wilson, G. (2009). Construct validity of the Wechsler abbreviated scale of intelligence and wide range intelligence test: Convergent and structural validity. School Psychology Quarterly, 24(4), 252–265. https://doi.org/10.1037/a0018030
