Overview of Outcome Measures Used Under the ACD

Thomas W. Frazier, PhD

May 6, 2024

Disclosures

I have received funding or research support from, acted as a consultant to, received travel support from, and/or received a speaker's honorarium from the PTEN Research Foundation, SYNGAP Research Fund, Malan Syndrome Foundation, ADNP Kids Research Foundation, Quadrant Biosciences, Autism Speaks, Impel NeuroPharma, F. Hoffmann-La Roche AG Pharmaceuticals, the Cole Family Research Fund, Simons Foundation, Ingalls Foundation, Forest Laboratories, Ecoeos, IntegraGen, Kugona LLC, Shire Development, Bristol-Myers Squibb, Roche Pharma, MaraBio, Scioto Biosciences, National Institutes of Health, and the Brain and Behavior Research Foundation, has equity options in Quadrant Biosciences/Autism Analytica, MaraBio, and Springtide, and has an investor stake in Autism EYES LLC, iSCAN-R, and AI-Measures.

Outline

- ACD Outcome Measures
  - Strengths
  - Weaknesses

- Recommendations for a Better Outcome Assessment Process

- **Major Caveat:** No assessment instrument is perfect. All assessment involves trade-offs. The questions are what tradeoffs should be made AND how do we optimize battery creation for the DoD TRICARE population given the current state of knowledge and available instruments

Current DoD / TRICARE Outcome Assessment Battery

Note: In interest of time, I cannot review every detail of the instruments and am only covering the major weaknesses applicable in this context

Vineland-3

Measure development recommendations from FDA, PROMIS, Boateng (2018) not followed with no or partial consideration in most areas

		Vineland-3
1	Stakeholder involvement from conceptualization to validation	Partial – prior administrators, clinicians, and clinical researchers
2	Use of qualitative research processes, including concept elicitation interviews, in determining domains/sub-domains guiding measurement coverage	Not Present
3	Creation of a domain map or disease model to guide measure creation	Not Present
4	Consideration of CIBI population-specific issues	Partial – IDD population
5	Attention to the nature of anticipated measure utilization (e.g., diagnostic, outcome measurement)	Not Present
6	Use of cognitive interviewing to evaluate item appropriateness	Partial – for teacher-report versions
7	Assessment of item relevance, clarity/readability, and potential bias	Present - Review for bias and relevance/ importance

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer. *Front Public Health, 6, 149. https://doi.org/10.3389/fpubh.2018.00149*

PROMIS® Validity Standards Committee on behalf of the PROMIS® Network. (2013). *PROMIS® instrument development and psychometric evaluation scientific standards.*

Frazier, T., Youngstrom, E., Frazier, A., Womack, R., & Uljarevic, M. (under review). A Critical Appraisal of Current Adaptive Behavior Measurement in Comprehensive Intensive Behavioral Intervention Outcome Assessment.

https://osf.io/preprints/psyarxiv/y549v

Vineland-3 cont.

Psychometric Characteristic	Vineland-3
Factor structure	Inconsistent findings across studies
Measurement invariance / DIF	Good for demographics, inconsistent for clinical groups
Measurement guides scoring	Not present
Scale reliability	Strong for scored scales
Conditional reliability	Not evaluated
Test-retest reproducibility	Very good
Test-retest reliability	Not evaluated
Content coverage	Good for broad domains, not adequate for modern understanding of social communication domains
Convergent validity	Strong evidence for broad domains
Discriminant validity	Excellent
Sensitivity to change	Strong evidence from clinical trials and observational studies, item scaling not created for tracking change during behavioral intervention

Psychometric characteristics are good in several traditional ways but inadequate in key ways with relevance to the DoD TRICARE application, particularly content and construct validity related to treatment planning and monitoring

Vineland-3 cont.

Norming / Scoring / Reporting	Vineland-3
Sample representativeness	Strong representativeness
Appropriate demographic adjustment	Yes, age adjusted standard scores
Norming type	Traditional with accounting for non-linear trends
Validity indicators	Not present
Raw scores	Present by domain and subscale
Norm-referenced scores	Present by domain and subscale
Growth scores	Present by domain
Attribute-adjusted change scores	No present
Reliable change	Not present
Clinically-meaningful change	Published for one application setting, not behavioral intervention

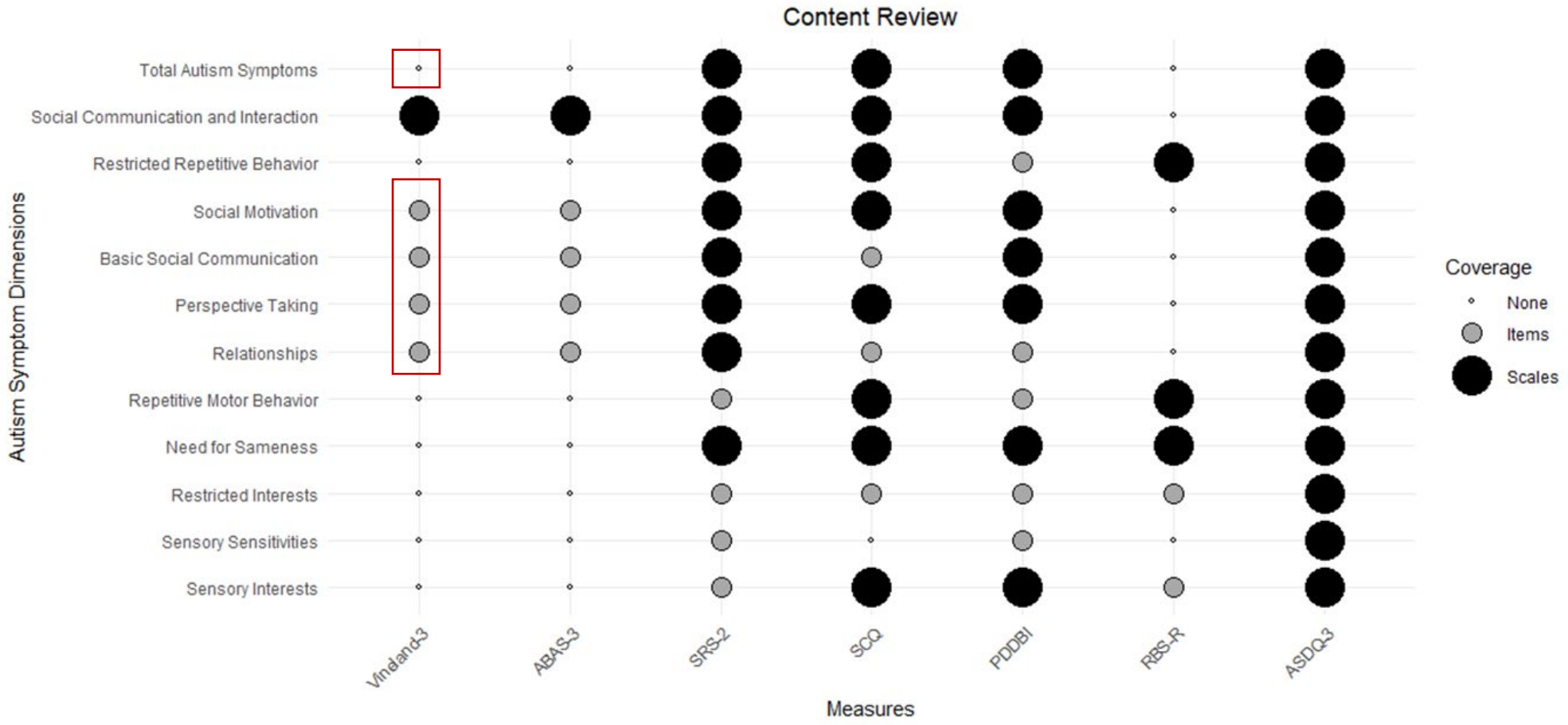
Scores are lacking for autism-relevant behavioral intervention treatment planning and progress monitoring. Lack of reliable change and attributed-adjusted change significantly limits utility clinical behavioral intervention

Vineland-3 cont.

Pragmatic Criteria	Vineland-3
Automated administration	Yes
Automated scoring	Yes
Visual display of cross-sectional results	Yes
Automated interpretive statements	Yes
Facilitated intervention target ID	Partial – item/target lists provided
Additional clinical guidance	Limited intervention strategy guidance
Automated progress monitoring	Very limited
Connection of intervention targets with teaching plans	Not present
Integration with other measures / domains to facilitate clinical workflow	Not present

Standard automation is present, consistent with legacy psychological assessment measures but lacking in intervention decision support, progress monitoring tools, or development of teaching plans

Vineland-3 cont.



Lack of social communication and interaction scales severely limits treatment planning and progress monitoring within a behavioral intervention context

SRS-2

Strengths

- Long history of clinical and research use
- Parent- and teacher-report versions
- Includes assessment of SCI and RRB symptoms
- Measures across a wide range of autistic traits
- Total score has excellent psychometric properties
- Has shown sensitivity to change
  - Although likely susceptible to rater and placebo effects, as are many parent-report questionnaires

Weaknesses

- Not built for ABA outcome assessment, wide range of stakeholders not involved in measure construction
- **Does not measure all ASD symptom sub-domains**
  - **RRB coverage is very weak**
- **Very limited utility for intervention planning**
- **Not all items are applicable to all ASD cases**
  - **require speech or imply higher levels of cognitive function**
- Test publisher maintains tight control of administration, with limited potential to automate administration and databasing of data with other measures

PDDBI

Strengths

- Parent and teacher forms
- Includes a wide range of core autism and associated symptoms
- Covers anxiety and aggression
- Covers receptive and expressive aspects of communication
- Substantial prior use in ASD intervention studies
- Content has utility for intervention planning

Weaknesses

- Published in 2003
  - Most published papers are not examining psychometric properties
- 30+ minute administration
- Not built for ABA outcome assessment
- Factor structure is not well understood
  - Unclear scoring
- Mixed evidence of sensitivity to change in behavioral intervention studies
- Which scores to use for tracking change?
- Test publisher maintains tight control of administration, with limited potential to automate administration and databasing of data with other measures

PSI-4 / SIPA

Strengths

- Long history of clinical and research use
- Massive literature, mostly outside autism
  - Google Scholar – 1470 results
- Total score has good reliability and construct validity evidence
- Strong coverage of content areas related to parent feelings of competence and stress, including child characteristics and parent-child relationship
- Prior studies in similar populations

Weaknesses

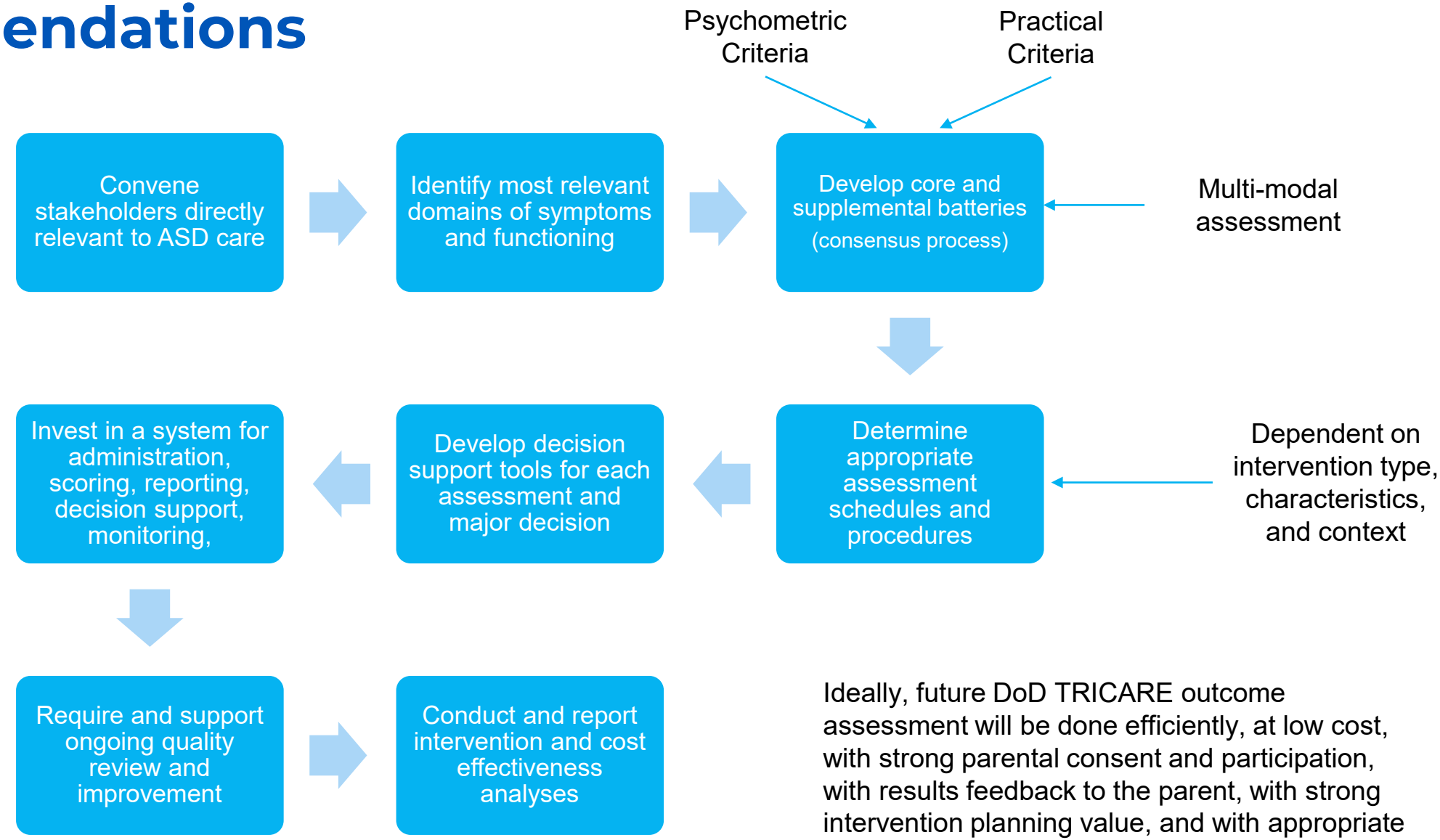
- 120 / 112 items
- 20-30 minute administration
- Significant differences in content across forms
- Not built for ABA outcome assessment, stakeholders not involved in measure construction
- Factor structure is not well replicated
  - Unclear scoring
- No evidence of sensitivity to change in behavioral intervention
- Test publisher maintains tight control of administration, with limited potential to automate administration and databasing of data with other measures
- Not consistently being used to inform treatment planning or provide resources, lack of caregiver feedback

Summary

ACD Battery has significant limitations that necessitate a revision in thinking about the optimal outcomes assessment approach

- ACD battery requirements are causing delays or gaps in services
- Not built for context, limited stakeholder input, and with mixed psychometric properties
- Does not adequately assess all of core autism sub-domains that have been identified
  - Social motivation, sensory sensitivity, sensory interests
- Limited coverage of co-occurring conditions relevant to child functioning
- Coverage of parent stress-related construct but lacking broader coverage of quality of life for the child/family/ecosystem
- Limited automation for progress monitoring, clinical decision support is very weak for intervention planning

Recommendations



Ideally, future DoD TRICARE outcome assessment will be done efficiently, at low cost, with strong parental consent and participation, with results feedback to the parent, with strong intervention planning value, and with appropriate supports to inform evidence-based clinical management decisions

THANK YOU

tfrazier@jcu.edu