

Portland State University

PDXScholar

School of Social Work Faculty Publications and Presentations

School of Social Work

2-2024

Measurement Matters: A commentary on the state of the science on patient reported outcome measures (PROMs) in autism research

Hillary Schiltz

University of California Los Angeles

Zachary J. Williams

Vanderbilt University School of Medicine

Shuting Zheng

Universtiy of California San Francisco

Elizabeth A. Kaplan-Kahn

Children's Hospital of Philadelphia

Hannah E. Morton

Oregon Health & Science University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/socwork_fac



Part of the [Social Work Commons](#)

Let us know how access to this document benefits you.

Citation Details

Schiltz, H. K., Williams, Z. J., Zheng, S., Kaplan-Kahn, E. A., Morton, H. E., Rosenau, K. A., ... & Autism PROMnet. (2023). Measurement matters: A commentary on the state of the science on patient reported outcome measures (PROMs) in autism research. *Autism Research*.






This Article is brought to you for free and open access. It has been accepted for inclusion in School of Social Work Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Authors

Hillary Schiltz, Zachary J. Williams, Shuting Zheng, Elizabeth A. Kaplan-Kahn, Hannah E. Morton, Kashia A. Rosenau, Christina Nicolaidis, Alexandra Sturm, and multiple additional authors

COMMENTARY

Measurement matters: A commentary on the state of the science on patient reported outcome measures (PROMs) in autism research

Hillary K. Schiltz¹ | Zachary J. Williams^{2,3,4,5,6}  | Shuting Zheng⁷  |
 Elizabeth A. Kaplan-Kahn⁸  | Hannah E. Morton⁹ | Kashia A. Rosenau¹⁰ |
 Christina Nicolaidis^{9,11} | Alexandra Sturm¹²  | Catherine Lord¹  | Autism PROMnet

¹Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, California, Los Angeles, USA

²Medical Scientist Training Program, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

³Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, Tennessee, USA

⁴Vanderbilt Brain Institute, Vanderbilt University, Nashville, Tennessee, USA

⁵Frist Center for Autism and Innovation, Vanderbilt University, Nashville, Tennessee, USA

⁶Vanderbilt Kennedy Center, Vanderbilt University Medical Center, Nashville, Tennessee, USA

⁷Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San Francisco, California, USA

⁸Center for Autism Research, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

⁹Department of Psychiatry, Oregon Health & Science University, Portland, Oregon, USA

¹⁰Department of Medicine, University of California Los Angeles, California, Los Angeles, USA

¹¹School of Social Work, Portland State University, Portland, Oregon, USA

¹²Department of Psychological Science, Loyola Marymount University, Los Angeles, California, USA

Correspondence

Hillary K. Schiltz, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, 760 Westwood Plaza, Los Angeles, CA 90024, USA.
 Email: hschiltz@mednet.ucla.edu

Abstract

High quality science relies upon psychometrically valid and reliable measurement, yet very few Patient Reported Outcome Measures (PROMs) have been developed or thoroughly validated for use with autistic individuals. The present commentary summarizes the current state of autism PROM science, based on discussion at the Special Interest Group (SIG) at the 2022 International Society for Autism Research (INSAR) Annual Meeting and collective expertise of the authors. First, we identify current issues in autism PROM research including content and construct operationalization, informant-structure, measure accessibility, and measure validation and generalization. We then enumerate barriers to conducting and disseminating this research, such as a lack of guidance, concerns regarding funding and time, lack of accessible training and professionals with psychometric skills, difficulties collecting large representative samples, and challenges with dissemination. Lastly, we offer future priorities and resources to improve PROMs in autism research including a need to continue to evaluate and develop PROMs for autistic people using robust methods, to prioritize diverse and representative samples, to expand the breadth of psychometric properties and techniques, and to consider developing field specific guidelines. We remain extremely optimistic about the future directions of this area of autism research. This work is well positioned to have an immense, positive impact on our scientific understanding of autism and the everyday lives of autistic people and their families.

Lay Summary

Much of autism research uses questionnaires completed by autistic people themselves or someone who knows them well. However, we often do not have a good understanding of how these types of questionnaires work among autistic people. In this commentary paper, we summarize important issues in this area of science, identify things that may get in the way of making progress on these issues, and provide recommendations and resources about what to do next.

KEYWORDS

autism research methods, autistic self-report, measurement, patient reported outcome measures, PROM, questionnaires

Hillary K. Schiltz and Zachary J. Williams contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *Autism Research* published by International Society for Autism Research and Wiley Periodicals LLC.

INTRODUCTION

High-quality science relies upon psychometrically valid and reliable measurements that are appropriate for the context of use (Flake & Fried, 2020; Fried et al., 2022). Within autism research, questionnaire data serves as the foundation of approximately 57% of all federally-funded research in the United States (Harris et al., 2021). Questionnaires are a type of patient-reported outcomes measure (PROM)—“any report of the status of a patient’s health condition, health behavior, or experiences(s) that comes directly from the patient” (Cella et al., 2015). Yet, very few PROMs have been developed or thoroughly validated for use with autistic people (Mazefsky, Day, et al., 2018; Scahill et al., 2019; Williams et al., 2021).

A growing body of empirical evidence indicates that issues of PROM psychometric investigation and validation are particularly relevant for autistic people due to (1) the unique topography of autistic experiences (e.g., differences in definitions of quality of life, anxiety experiences, disability-specific peer victimization) that are not captured by PROMs that were created and validated in the general population and therefore require focused investigation; (2) the traditional reliance within the field on proxy-report due to characteristics of autism itself (e.g., differences in communication) and the emerging realization that discrepancies with self-reported autistic perspectives reflect distinct experiences; (3) the inherent gender imbalance of boys and men within the autistic population that leads to biased samples; and (4) compounding of systemic access issues for those from multiply marginalized groups (i.e., autistic people with other historically marginalized identities).

Therefore, researchers have called for increased attention and efforts toward improving measurement in autism research through methodologically rigorous PROM development, evaluation, and adaptation (e.g., Bal et al., 2018; Nicolaidis et al., 2020; Wagner et al., 2020). In response, we launched a Special Interest Group (SIG) at the 2022 International Society for Autism Research (INSAR) Annual Meeting focused on PROMs and autism.¹ Here, we summarize the discussion among stakeholders, researchers, and clinicians during the 2022 INSAR Annual Meeting SIG Session, and expand these perspectives to include the collective expertise of the authors to identify current issues in autism PROM research, enumerate barriers to conducting and disseminating this research, and offer future recommendations and resources to strengthen methodological approaches within the autism research community and improve PROMs in autism research.

¹SIG Title: Improving Patient Reported Outcome Measures (PROMs) in Autism Research: Bridging the Gap Between Advanced Psychometric Techniques and Stakeholder Priorities. Additional information about the SIG, as well as discussion questions, agenda, and other SIG session materials can be found at: <https://www.autism-insar.org/page/SIG2022PROMs>.

Author positionality and SIG members

We believe it is important to acknowledge the positionality of the authors who contributed to this commentary and the members of the SIG. The author team includes nine individuals, all of whom have extensive research and clinical experience working with autistic individuals. Authors have backgrounds in the following disciplines: clinical psychology (HKS, HEM, AS, CL), developmental psychology (KAR), internal medicine (CN), psychiatry/neuroscience (ZJW), school psychology (EAK-K), and special education/behavior analysis (SZ). Three authors identify as neurodivergent (two as both autistic/ADHD; one as non-autistic but ADHD) and the remainder identify as neurotypical. Eight of the authors are cisgender women and one is a cisgender man. Eight of the authors identify as White and one as Asian, with one author identifying as Hispanic and the remaining eight identifying as non-Hispanic. Lastly, multiple authors identify as applied psychometricians. The SIG is composed of members across many stages of training (i.e., graduate student, postdoctoral scholar, research scientist, faculty) and professional settings (e.g., academia, industry, non-profit). Multiple SIG members also have lived experiences relevant to the current topic (e.g., as autistic adults or caregivers of autistic individuals).

CURRENT ISSUES IN AUTISM PROM RESEARCH

The dearth of available PROMs with strong psychometric properties in the autistic population is rooted in current issues regarding operationalization and content, informant-structure, measure accessibility, and measure validation (Table 1).

Content and construct operationalization

There are concerns that some of the PROMs used in autism research may not have adequate content validity; that is, PROMs may not assess all aspects of the construct of interest or may capture irrelevant constructs. Furthermore, some constructs within autism research are conceptually ambiguous or do not have an agreed upon operational definition (e.g., controversial constructs). Because of these issues, PROM item generation is often not guided by a pre-specified theoretical model or a clear, operationalized definition of the construct of interest (e.g., using Patient-Reported Outcome Measurement Information System (PROMIS²) method; PROMIS Health Organization, 2013; see also (Mazefsky, Day,

²The PROMIS method was developed by an NIH Roadmap Initiative with leading experts in the field to create tools that assess a wide range of health constructs (e.g., fatigue, emotional distress, pain) efficiently and with excellent precision. Briefly, the PROMIS method includes a set instrument development and validation standards that provide the scientific foundation for the PROMIS instruments including defining the concept and conceptual model, composing the item pool, and testing instrument properties, among others.

et al., 2018). For example, social “camouflaging” (Cook et al., 2021) recently gained substantial research attention, sparking debates about construct validity and clinical interpretation (Fombonne, 2020; Lai et al., 2021; Williams, 2022) that are critical for ensuring sound measure development and psychometric validation (Hannon et al., 2023; Williams, 2022). Additionally, overlapping constructs pose an issue. That is, it may be difficult to distinguish between an instrument’s target construct (e.g., anxiety) and autistic traits or related behaviors that could be but are not necessarily related to that construct (e.g., repetitive behavior or fidgeting). For example, measures intended to capture autism traits may be elevated among children with more externalizing behavior and lower language and cognitive skills (Hus et al., 2013) and that responses on measures of anxiety may be impacted by autism characteristics (Schiltz & Magnus, 2021). As another example, an item on a sensory reactivity questionnaire about avoiding noisy parties may be influenced by the respondent’s social motivation or social anxiety. Similarly, constructs may not fully capture all relevant content areas that are important to autistic people. For example, existing measures of quality of life may not adequately incorporate the importance of a positive autistic identity, sensory issues, or a sense of contribution to society (McConachie et al., 2020). Finally, constructs may be conceptualized in ableist ways, for example, confusing health-related quality of life with underlying disability or assuming that a high degree of social contact or relatedness are necessary components of self-determination or emotional well-being.

Informant structure

There are also concerns that suitable self-report instruments are not always available in autism research (Nicolaidis et al., 2020), which necessitates gathering information via informant-report from other informants (e.g., caregivers, teachers). While informant report can provide relevant and helpful information, it is problematic to assume that self- and informant-ratings represent the same underlying construct and are interchangeable. Indeed, there are meaningful differences between self- and informant-reports for a wide variety of PROMs (De Los Reyes, 2011), for example in the area of self-determination (Shogren et al., 2021; Tomaszewski et al., 2020), and these discrepancies may be moderated by demographic and clinical characteristics of both raters (Taylor et al., 2022). Self-report is also particularly important for subjective constructs such as mood and well-being, as well as behaviors that may not be observable by informants (e.g., restricted and/or repetitive behaviors and interests (RRBs) or bullying at school during unstructured time; Demaray et al., 2013; McDermott et al., 2020). Nevertheless, some constructs (e.g., those representing rigid personality features, pragmatic language impairments, or social aloofness (Sasson et al., 2014)) may potentially be more accurately

TABLE 1 Current issues in autism patient reported outcome measure (PROM) research.

Areas of concern	Specific issues
Content and construct operationalization	Constructs not yet clearly operationalized or conceptually ambiguous
	Lack of agreement on operationalization (e.g., controversial constructs)
	Mismatch between measure content and theoretical construct definition
	Many relevant constructs poorly captured by existing PROMs (e.g., capturing irrelevant constructs, ableist conceptualizations of constructs, lack of coverage of relevant subconstructs identified by theory, clinical experts, and/or autistic lived experiences)
Informant-structure	Over-reliance on informant-report and under use of self-report
	Assumption that self- and informant-report measure same underlying construct
Measure accessibility	Problems with item wording (e.g., complicated terms/phrases, figures of speech, jargon)
	Confusing response options (missing interval, leading question)
Measure validation and generalization	Limited psychometric data in the autistic population and/or relevant subgroups (e.g. adults, individuals with below average cognitive abilities, gender minorities), especially beyond initial studies
	Some psychometric properties are less often tested including sensitivity to change, cross-cultural validity, test-retest reliability, measurement invariance, or predictive validity. Factor structure, internal consistency, and concurrent validity are more often examined.
	Outdated methods used to assess latent structure, reliability, construct validity

represented by informant as opposed to self-ratings due to limited insight on the part of the individual being rated. Although self-report should be included when possible, we note that there are indeed some populations (e.g., very young children, individuals with severe/profound intellectual disability) who may not be able to complete self-report questionnaires, but autism per se does not preclude the use of self-report. Utilizing participatory research methods (Keating, 2021; Poulsen et al., 2023) that involve of autistic people and related stakeholders in PROM development and/or use of other complimentary measures (e.g., metrics of reading level, language ability, alexithymia) can be particularly helpful in determining the proper informant structure for a given construct or research question (e.g., in what contexts self and caregiver report are appropriate).

Measure accessibility

Concerns with PROMs in autism research often pertain to issues of accessibility; that is, common issues in developing PROM items and responses (which are not unique to autism research) may be magnified for autistic people. For example, items may have problematic wording such as complicated phrases (e.g., “to be more ineffective”), technical jargon (e.g., “engage in social interaction”), difficult vocabulary (e.g., “self-determination”), and figures of speech (e.g., “going your way”) (Choi & Pak, 2004; Nicolaidis et al., 2020; Zickar, 2020). Furthermore, response options may be confusing due to lack of clarity and/or precision (e.g., What proportion of time is “some of the time?” What does “moderately bothered” mean?; Nicolaidis et al., 2020). In addition to wording issues, detail-oriented autistic people may feel worried or unsure how to respond to items with 100% accuracy (Nicolaidis et al., 2020). Collectively, these problems impede accessibility of PROMs for many autistic people, including individuals *without* intellectual disability or literacy challenges.

Measure validation and generalization

Even for PROMs that have been developed or adapted for autistic people, there is often limited psychometric data on the use of these measures in the autistic population or relevant subgroups of autistic people, especially beyond initial validation studies. This consideration is crucial given that psychometric properties of a measure may not generalize outside the validation sample, especially for individuals from historically underrepresented groups such as racial and gender minorities. Scoring of measures also often requires a binary gender selection, which raises additional validity concerns given the disproportionate gender variance within the autistic community (Pecora et al., 2020). Furthermore, most psychometric studies of PROMs in autism have been conducted with youth and people with average or above cognitive abilities (for a review, see Kim & Lecavalier, 2022) and PROMs are often not developed with consideration of autistic adults, those with intellectual disability (particularly in the moderate or severe/profound range), or those who do not communicate with spoken words. Notably, issues of validation and generalization are certainly not exclusive to autism research, and there is ongoing debate in the literature regarding the expectation for full psychometric validation in the target population (Hughes, 2018). Nonetheless, the immense heterogeneity among autistic people, especially in terms of cognitive and language levels, at a minimum necessitates evaluation of PROMs across samples that vary on these characteristics (or explicit restriction of PROM scope to only a subset of autistic individuals for whom they are expected to be appropriate).

Finally, the majority of PROMs in autism research are psychometrically tested in cross-sectional studies, and often only include assessment of factor structure, internal consistency, and concurrent validity, many with outdated statistical methods (such as those critiqued by Preacher & MacCallum, 2003 two decades ago). This approach limits the types of validity data available to support these measures. Fewer studies test sensitivity to change, cross-cultural validity, test–retest reliability, measurement invariance, or predictive validity.

PROGRESS ON PROMS IN AUTISM RESEARCH

Fortunately, a growing number of researchers are working to improve PROMs in autism. For example, some investigators have developed PROMs in partnership with autism stakeholders or conducted cognitive interviews with potential respondents, greatly increasing the overall content validity of the novel measures and their applicability to the autistic population (e.g., Cassidy et al., 2018; Mazefsky, Day, et al., 2018; Nicolaidis et al., 2021; Riccio et al., 2020). Others have used a community based participatory research approach with autistic adults to adapt existing instruments for use with autistic participants (Nicolaidis et al., 2020). Other approaches include the application of methods such as item response theory (IRT), structural equation modeling, and network analysis, and even more standard approaches such as test–retest and convergent validity to answer substantive empirical questions about the psychometric properties of both newly-created and existing PROMs in the autistic population (e.g., Frazier et al., 2022; Mazefsky, Yu, et al., 2018; Schiltz et al., 2019; Schiltz & Magnus, 2020, 2021; Uljarević et al., 2022; Williams et al., 2021; Williams, Cascio, & Woynaroski, 2023). Still others have begun to incorporate PROMs into routine clinical practice, and the first systems of *measurement-based care* (Lewis et al., 2018) were recently described within autism research (McFayden et al., 2021; Schwartzman et al., 2023).

The above efforts have resulted in notable progress on PROMs in autism. Available measures of anxiety with sound psychometric evidence in autism have grown substantially within the past 10 years (Schiltz et al., *in press*). For example, the Anxiety Scale for Children with Autism Spectrum Disorder (ASC-ASD; Rodgers et al., 2016) and upward extension the Anxiety Scale for Autism-Adults (ASA-A; Rodgers et al., 2020) were developed with input from key stakeholders, include items based on the research evidence on anxiety in autism (e.g., sensory processing, intolerance of uncertainty), and received psychometric evaluation in samples of autistic people. As another example, the construct of emotion dysregulation, a common concern among autistic people (Cai

et al., 2018; Mazefsky et al., 2013), can be measured with the Emotion Dysregulation Inventory, an informant questionnaire created using PROMIS methods, which involved stakeholder input, cognitive interviews, a pre-specified conceptual model, and IRT analyses in both general population samples and autistic samples (Mazefsky et al., 2021; Mazefsky, Day, et al., 2018; Mazefsky, Yu, et al., 2018). As a final example, the Beck Depression Inventory-II (Beck et al., 1996) was rigorously psychometrically evaluated in autistic and general-population samples using IRT methods, with differential item functioning tests establishing approximate equivalence between the two populations and a free online score calculator provided to allow for easy calculation of IRT-based latent trait scores in research and clinical practice (Williams et al., 2021). Notably, these are only a few examples of extensive ongoing research to improve PROMs in autism research.

CURRENT BARRIERS TO AUTISM PROM RESEARCH

Despite the clear need for continued efforts to develop, adapt, and evaluate PROMs for autistic people, there are barriers to engaging in and disseminating this research.

Uncertainty about where to start and where to go

For researchers aiming to improve PROMs for use in the autistic population, it is often unclear if it would be more appropriate to adapt an existing questionnaire or develop an entirely new measure. Once a decision is made either adapt or develop a measure, generating appropriate items can also be difficult as there is no systematic process for item-writing, curation, and evaluation (DeWalt et al., 2007).

Concerns regarding funding and time

Desire to advance PROM research in autism intersects with concern that funders have little interest in “basic measurement” research, particularly evaluating the properties of existing as opposed to newly-developed measures. Adequate funding support is needed to allow researchers to devote time and effort to conducting measurement studies.

Lack of accessible training and professionals with psychometric skills

The ability to carry out psychometric analyses, or collaborate with someone who can, is also necessary for researchers wishing to evaluate PROMs in autism. While

psychometrics at a basic level is typically part of standard graduate school curriculum in psychology and allied disciplines, more advanced methods necessitate more advanced training. Thus, researchers wishing to go beyond the basics must seek out external resources and training. On the other hand, for autism researchers who are not interested in conducting psychometric research themselves, foundational psychometric knowledge is still necessary to make optimal methodological decisions and interpretations regarding PROMs (Flake & Fried, 2020).

Difficulties collecting large representative samples

Large samples (i.e., in the high hundreds or thousands) are often required to conduct rigorous psychometric evaluations of measure (e.g., using IRT). It can be difficult to obtain autistic samples of this size, related to the funding and time barriers described above. Although large databases (e.g., Feliciano et al., 2018; Hall et al., 2012) are promising and likely an important step for future research, straightforward use of these samples for PROM evaluation is often precluded by a lack of clarity in sample characterization, questionable representativeness of large samples (Rødgaard et al., 2022), time-intensive data cleaning, and fraudulent responses within online surveys (Harrop et al., 2021). The pooling together of many smaller samples that utilize the same or similar measures in an integrative data analysis is one promising alternative that has previously been leveraged within the autism literature (e.g., Magiati et al., 2017; Sturm et al., 2017; Williams, Schaaf, et al., 2023).

Challenges with dissemination

It can also be challenging to disseminate research on PROMs validated in autistic samples (and subsequent measure updates) with the wider autism community. There is a perceived lack of interest in conference presentations that focus on measure development, and few peer-reviewed journals specifically address measurement issues in autism or related areas. Dissemination of and access to licensed PROMs are further limited by challenges related to obtaining license permission, including cost and administrative barriers (e.g., cumbersome contract processing). It is a difficult balance to protect the intellectual property of measure developers and to make measures easily accessible to a larger audience (Newman & Feldman, 2011). It is also often a challenge to determine *which* PROM(s) to select, even when multiple reliable and valid PROMs are available for use with autistic samples. Taken together, these limitations likely contribute to underutilization and inconsistency of valid PROM use across autism research studies, which further

hinders the ability to compile large datasets (as described above). On the other hand, open science (McKiernan et al., 2016) offers a platform that is increasingly used to make PROMs publicly accessible for research and clinical use (e.g., PROMIS measures are freely available online).

FUTURE PRIORITIES FOR PROMS IN AUTISM RESEARCH

Below, we identify five priorities for future research to advance the field of PROM science in autism (Table 2).

Continue to evaluate and develop PROMs for autism

There is a need for continued research efforts that develop and evaluate PROMs specific to autism. Although there is understandable hesitancy to apply for funding to support such efforts (as described above), there are also funding exemplars for this type of research (e.g., NIMH: Autism Patient Reported Outcomes Measures [AutPROM] Toolbox [PI: Nicolaidis]; NICHD: Emotion Dysregulation [PI: Mazefsky]; AIR-P: Quality of Life PROM [PIs: Lory & Maddox]; Flourishing PROM [PI: Ross]; Autism Speaks, Loneliness PROM [PI: Schiltz]; PCORI: multiple PROMs studies, not

autism-specific). We hope the growing recognition of the importance of PROM science in autism leads to an even greater emphasis on measure evaluation, adaptation, and development by researchers and funding organizations.

It also remains paramount that researchers determine whether existing PROMs developed for use in the general population or other clinical groups can be suitably applied to autistic groups without modification/adaptation. In some instances, minor modifications may be warranted. In these instances, researchers should also prioritize reducing participant burden and fatigue (e.g., by developing shortened or computerized adaptive test versions; Lyall et al., 2022). However, when PROMs are determined to be unsuitable due to psychometric limitations, we recommend designing and validating novel PROMs specifically for autistic individuals.

Use robust methods to develop PROMs

We recommend utilizing a systematic approach when developing new PROMs, similar to the PROMIS framework, to address many of the aforementioned issues in PROM science in autism. These steps include (1) construct definition and operationalization; (2) item bank development; (3) qualitative item review; (4) pilot study; (5) dimensionality analysis; (6) evaluation of measurement invariance/differential item functioning; and (7) calibration and scoring. Importantly, stakeholders should be

TABLE 2 Future priorities for autism patient reported outcome measure (PROM) research.

Priorities	Specific recommendations
Continue to evaluate and develop PROMs for autism	<p>More studies examining whether PROMs developed in other populations (e.g., general population, other clinical groups) can be suitably applied to autism</p> <p>When minor modifications are not possible, design and validation of novel PROMs specifically for use in autistic individuals and related clinical populations</p> <p>Increased efforts to reduce participant burden for new, existing, and adapted PROMS (e.g., short versions or computerized adaptive test versions)</p>
Use robust methods to develop PROMs	<p>More participatory measure development and validation studies</p> <p>Thorough conceptual mapping of construct of interest</p> <p>Cognitive interviewing incorporated into the content validation process</p>
Prioritize diverse and representative samples	<p>Assessment of psychometric properties in multiple subpopulations defined by demographic and clinical factors (particularly among autistic women, adults, individuals with below average cognitive abilities, and people who are non-speaking)</p> <p>Periodically reassess the psychometric properties of PROMs with people who meet current diagnostic conceptualization</p> <p>Coordination of measure collection across studies of well-characterized autistic samples and subsequent data sharing</p>
Expand breadth of psychometric properties and techniques	<p>Longitudinal studies leveraged for psychometric information that requires multiple timepoints (e.g., test-retest reliability, sensitivity to change, temporal measurement invariance)</p> <p>Leveraging of recent statistical advances (e.g., exploratory graph analysis)</p> <p>Psychometric work should be undertaken by research groups who did not create the measure(s) being evaluated</p>
Consider developing field specific guidelines	<p>Guidelines for selecting PROMs in autism research (general and construct specific)</p> <p>Guidelines for conducting and reporting upon psychometric properties of PROMs in autism research</p>

involved throughout the PROM development process (i.e., defining the construct, item development, review of items and response options, etc.). There is a need for theoretically driven item development based on multiple information sources including stakeholder perspectives and systematic review of previous research. Qualitative item review should be conducted by academic/clinical content experts *and* autistic stakeholders (e.g., McConachie et al., 2020). Cognitive interviews with intended respondents during qualitative item review (e.g., Ratto et al., 2022) are also critical to determine if items are measuring intended constructs in an accessible way. These interviews can then guide the development of more accessible self-report measurement methods (e.g., incorporating pictorial scales, plain language, and/or assistance from a support person; Fitzpatrick et al., 2022; Kunin, 1955; Sauer et al., 2020).

Prioritize diverse and representative samples

Psychometric properties of both novel *and* existing PROMs should be examined in subpopulations of autistic individuals defined by demographic and clinical factors (e.g., comparing autistic people across gender, individuals with and without cognitive impairments, speaking and non-speaking individuals). One approach for collecting large, diverse, and representative samples involves the coordination of measure collection across studies of well-characterized autistic samples and subsequent data sharing. Another approach is to concomitantly use multiple different (or complementary) recruitment strategies (e.g., recruiting from a university autism clinic and online via social media), so as to counterbalance limitations in each one. Given that individuals who identify as cisgender, White males are over-represented in existing clinical/community samples, and efforts to establish more representative samples are needed. Researchers should not assume that the highly-selected samples of participants willing and able to take part in traditional autism research studies (or conversely surveys that predominantly recruit from social media or large-scale participant pools such as SPARK) are representative of the entire autistic population. As such, studies should test PROMs in naturalistic environments; should use eligibility criteria that match what would be used in real-world settings (e.g., eligibility for services as opposed to a researcher-based assessment of autism diagnosis); should provide adequate accommodations and supports; should partner with organizations and leaders that can increase trust; and should actively work to lower barriers to participation. Language translation and cross-cultural comparison of measures, a common later step of the PROM development process, should also be undertaken in accordance with best practices (Eremenco et al., 2005).

Expand breadth of psychometric properties and techniques

Additional work in autism research should also focus on improving the breadth of psychometric validation studies in the field. There is a great need to perform more psychometric analyses on PROM data from multiple timepoints (e.g., test–retest reliability, sensitivity to change, and temporal measurement invariance), particularly in the context of secondary data analyses of large-scale longitudinal cohorts. Given the evolving landscape of psychometric methods (Jebb et al., 2021; Stover et al., 2019), researchers should also leverage recent statistical advances (e.g., exploratory graph analysis; Golino et al., 2022) to test psychometric properties. Ideally, this psychometric work should be conducted by research groups who did not create the measure(s) being evaluated to limit vested interests (e.g., financial, academic/reputational) in the PROM's validity or utility. Notably, many of the more advanced statistical approaches indeed require large samples, which may not be feasible given constraints on resources or the specific autistic sub-population under study (e.g., childhood disintegrative disorder, autism associated with rare neurogenetic syndromes). In such cases, rather than attempt to utilize “large-sample” statistical techniques, researchers are advised to use methods appropriate for the sample size, which may include classical test theory approaches (DeVellis, 2006; Nolte et al., 2019), Rasch models (Cleanthous et al., 2019; Petrillo et al., 2015), or Bayesian structural equation modeling (Smid et al., 2020; Ulitzsch et al., 2023) to analyze available data.

Continued validation of PROMs is also critical, as “validation is an ongoing process” (Chan, 2014) and no PROM should ever be simply considered “fully validated” for use in autism research after some specified amount of psychometric evidence has been accrued (Clark & Watson, 2019). Continuous validation may include periodic measure updates, for example to reflect current diagnostic conceptualizations of autism and the evolving population of individuals receiving the diagnostic label of “autism spectrum disorder” over time (Arvidsson et al., 2018).

Consider developing field specific guidelines

Best practice guidelines may be helpful for autism research in terms of PROM selection (both general and construct specific) and adaptation, as well as conducting and reporting on the psychometric properties of PROMs in the autistic population (i.e., an autism-specific extension of the CONsensus-based Standards for the selection of health Measurement INSTRUMENTS (COSMIN)

guidelines³; (Mokkink et al., 2010). Notably, a number of tools such as the COSMIN extensions (Gagnier et al., 2021; Mokkink et al., 2016, 2018, 2020; Prinsen et al., 2016, 2018; Terwee et al., 2018) and the ConPsy checklist (Vitoratou et al., 2024) are also available to guide the development and evaluation of psychometric instruments in general and can already be applied to PROMs within autism research in their current form. Use of measure-selection heuristics (e.g., selecting a PROM based on frequency of use or robustness of psychometric data in autistic people) may be beneficial under some circumstances. It remains an open question whether there is a need for guidelines specific to autism research (e.g., including autism-relevant aspects of the development process such as invariance testing by diagnostic status, assessment of confounding by intellectual functioning, or autistic stakeholder involvement) or if other non-specific guidelines are suitable.

RECOMMENDATIONS FOR ALL AUTISM RESEARCHERS

Although the above recommendations are important for autism researchers with a particular focus on PROM science, all autism researchers can help to improve measurement practices within the broader field of autism research. In particular, we recommend including the following elements for each PROM within the measures section of the method for all empirical articles (or, if not possible, as a supplement):

- Details regarding available psychometric properties in samples of autistic people. Relevant properties to report include reliability, validity, factor structure/dimensionality, measurement invariance/differential item functioning [particularly across diagnostic groups], diagnostic/screening performance (if the measure is intended for that purpose), and sensitivity to change. If this information is unavailable, as the PROM has not received psychometric evaluation in autism, this should be stated explicitly, and psychometric evidence from non-autistic populations should be reported with the caveat that such evidence may not directly transfer to the autistic population.
- A single-test index of internal consistency reliability within the study sample (preferably a version of coefficient omega [McNeish, 2018; Revelle & Condon, 2019], but if not feasible, Cronbach's coefficient alpha is suitable [Raykov & Marcoulides, 2019]).

³The mission of COSMIN is "to improve the selection of outcome measurement instruments of health outcomes by developing and encouraging the use of transparent methodology and practical tools for selecting the most suitable outcome measurement instrument in research and clinical practice." COSMIN has multiple useful tools including a detailed user manual outlining the 10-step procedure for conducting a systematic review on PROMs, risk of bias checklist, taxonomy of measurement properties, and database for systematic reviews, among others.

- If the study sample includes multiple test administrations (e.g., longitudinal studies or studies in which multiple informants fill out the same measure), an index of test-retest or inter-rater reliability (preferably the appropriate intraclass correlation [McGraw & Wong, 1996; ten Hove et al., 2022]).
- Details regarding any PROM modifications or adaptations in the present study.

RESOURCES FOR PROMs IN AUTISM RESEARCH

We describe ongoing initiatives and resources that we hope will help to address the many identified barriers to progress in this area of autism research.

Psychometric training resources

To help address the training barriers described above, we created the Autism PROMnet Training Resources List (<https://www.autismpromnet.org/view-training-resources.html>), a crowdsourced effort for researchers seeking resources and training related to psychometrics and measure development. These resources cover topics such as factor analysis, classical test theory, IRT, measurement invariance, measure development, and specific software and coding resources, among others. We invite the submission of psychometric training resources at <https://www.autismpromnet.org/submit-a-new-prom-to-repository.html>.

PROM dissemination and collaboration resources

There are several clear needs in the area of autism PROMs, including disseminating information to support researchers and clinicians in making informed PROM selections, as well as facilitating collaborations between psychometricians and quantitative methodologists who can lend their statistical expertise to autism research studies. In this context, we created an Autism PROMnet Measures Repository (<https://www.autismpromnet.org/visit-prom-repository.html>) as a resource for researchers and clinicians seeking PROMs with evidence for reliability and validity in samples of autistic people. The repository includes PROMs related to many topics (e.g., autism traits, mental health, quality of life) and is searchable by various categories including by author, year, and construct. This repository is a crowdsourced effort and provides a centralized location for PROM dissemination. We invite the submission of PROMs with psychometric evidence in autism at <https://www.autismpromnet.org/submit-a-training-resource.html>. We also created the Autism PROMnet Google Group (<https://groups.google>).

com/g/autism-promnet/), which serves as a general list-serv for researchers, clinicians, and autism stakeholders to connect and discuss PROMs, psychometrics, and measurement development in autism research.

CONCLUSION

Although we identify many issues with the current state of PROM science in autism research and acknowledge many barriers to progress, we are excited by the burgeoning group of people who share a passion for this topic. We remain extremely optimistic about the future directions of this area of autism research, and particularly in the context of the formation of Autism PROMnet as an organizing body and the development of many free and crowdsourced resources to improve autism measurement and PROM science. This work is well positioned to have an immense, positive impact on our scientific understanding of autism and the everyday lives of autistic people and their families.

ACKNOWLEDGMENT

The authors have nothing to report.

CONFLICT OF INTEREST STATEMENT

HKS, ZJW, SZ, EAK, HEM, and KR serve on the leadership team of Autism PROMnet, for which they receive no royalties or compensation. ZJW has received consulting fees from Roche and Autism Speaks. He is also the vice-chair of the Autism Intervention Network on Physical Health (AIR-P) ANSWER committee. All other authors have no conflicts of interest to declare that are relevant to the content of this article.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Zachary J. Williams  <https://orcid.org/0000-0001-7646-423X>

Shuting Zheng  <https://orcid.org/0000-0002-0821-2716>

Elizabeth A. Kaplan-Kahn  <https://orcid.org/0000-0002-2481-3247>

Alexandra Sturm  <https://orcid.org/0000-0002-5060-1838>

Catherine Lord  <https://orcid.org/0000-0001-5633-1253>

REFERENCES

- Arvidsson, O., Gillberg, C., Lichtenstein, P., & Lundström, S. (2018). Secular changes in the symptom level of clinically diagnosed autism. *Journal of Child Psychology and Psychiatry*, 59(7), 744–751. <https://doi.org/10.1111/jcpp.12864>
- Bal, V. H., Hendren, R. L., Charman, T., Abbeduto, L., Kasari, C., Klinger, L. G., Ence, W., Glavin, T., Lyons, G., & Rosenberg, E. (2018). Considerations from the 2017 IMFAR preconference on measuring meaningful outcomes from school-age to adulthood. *Autism Research*, 11(11), 1446–1454. <https://doi.org/10.1002/aur.2034>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory*. Harcourt Brace Jovanovich.
- Cai, R. Y., Richdale, A. L., Uljarević, M., Dissanayake, C., & Samson, A. C. (2018). Emotion regulation in autism spectrum disorder: Where we are and where we need to go. *Autism Research*, 11(7), 962–978. <https://doi.org/10.1002/aur.1968>
- Cassidy, S., Bradley, L., Bowen, E., Wigham, S., & Rodgers, J. (2018). Measurement properties of tools used to assess depression in adults with and without autism spectrum conditions: A systematic review. *Autism Research*, 11(5), 738–754. <https://doi.org/10.1002/aur.1922>
- Cella, D., Hahn, E. A., Jensen, S. E., Butt, Z., Nowinski, C. J., Rothrock, N., & Lohr, K. N. (2015). *Patient-Reported Outcomes in Performance Measurement*. RTI Press. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK424379/>
- Chan, E. K. H. (2014). Standards and guidelines for validation practices: Development and evaluation of measurement instruments. *Validity and Validation in Social, Behavioral, and Health Sciences*, 54, 9–24. https://doi.org/10.1007/978-3-319-07794-9_2
- Choi, B. C. K., & Pak, A. W. P. (2004). A catalog of biases in questionnaires. *Preventing chronic disease*, 2(1), A13.
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. In *Psychological assessment* (Vol. 31, pp. 1412–1427). American Psychological Association. <https://doi.org/10.1037/pas0000626>
- Cleanthous, S., Barbic, S. P., Smith, S., & Regnault, A. (2019). Psychometric performance of the PROMIS® depression item bank: A comparison of the 28- and 51-item versions using Rasch measurement theory. *Journal of Patient-Reported Outcomes*, 3(1), 47. <https://doi.org/10.1186/s41687-019-0131-4>
- Cook, J., Hull, L., Crane, L., & Mandy, W. (2021). Camouflaging in autism: A systematic review. *Clinical Psychology Review*, 89, 102080. <https://doi.org/10.1016/j.cpr.2021.102080>
- De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, 40(1), 1–9. <https://doi.org/10.1080/15374416.2011.533405>
- Demaray, M. K., Malecki, C. K., Secord, S. M., & Lyell, K. M. (2013). Agreement among students', teachers', and parents' perceptions of victimization by bullying. *Children and Youth Services Review*, 35(12), 2091–2100. <https://doi.org/10.1016/j.childyouth.2013.10.018>
- DeVellis, R. F. (2006). Classical test theory. *Medical Care*, 44(11), S50–S59. <https://doi.org/10.1097/01.mlr.0000245426.10853.30>
- DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, 45(suppl. 1), S12–S21. <https://doi.org/10.1097/01.mlr.0000254567.79743.e2>
- Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions*, 28(2), 212–232. <https://doi.org/10.1177/0163278705275342>
- Feliciano, P., Daniels, A. M., Snyder, L. G., Beaumont, A., Camba, A., Esler, A., Gulsrud, A. G., Mason, A., Gutierrez, A., & Nicholson, A. (2018). SPARK: A US cohort of 50,000 families to accelerate autism research. *Neuron*, 97(3), 488–493. <https://doi.org/10.1016/j.neuron.2018.01.015>
- Fitzpatrick, R., McGuire, B. E., & Lydon, H. K. (2022). Improving pain-related communication in children with autism spectrum disorder and intellectual disability. *Paediatric and Neonatal Pain*, 4(1), 22–32. <https://doi.org/10.1002/pne2.12076>
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465. <https://doi.org/10.1177/2515245920952393>

- Fombonne, E. (2020). Camouflage and autism. *Journal of Child Psychology and Psychiatry*, 61(7), 735–738. <https://doi.org/10.1111/jcpp.13296>
- Frazier, T. W., Khaliq, I., Scullin, K., Uljarevic, M., Shih, A., & Karpur, A. (2022). Development and psychometric evaluation of the Open-Source Challenging Behavior Scale (OS-CBS). *Journal of Autism and Developmental Disorders*, 53, 4655–4670. <https://doi.org/10.1007/s10803-022-05750-5>
- Fried, E. I., Flake, J. K., & Robinaugh, D. J. (2022). Revisiting the theoretical and methodological foundations of depression measurement. *Nature Reviews Psychology*, 1(6), 358–368. <https://doi.org/10.1038/s44159-022-00050-2>
- Gagnier, J. J., Lai, J., Mokkink, L. B., & Terwee, C. B. (2021). COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research*, 30(8), 2197–2218. <https://doi.org/10.1007/s11136-021-02822-4>
- Golino, H., Christensen, A. P., & Garrido, L. E. (2022). Invited commentary: Exploratory graph analysis in context. *Revista Psicologia: Teoria e Prática*, 24(3), ePTPPA14197. <https://doi.org/10.5935/1980-6906/ePTPIC15531.en>
- Hall, D., Huerta, M. F., McAuliffe, M. J., & Farber, G. K. (2012). Sharing heterogeneous data: The national database for autism research. *Neuroinformatics*, 10, 331–339. <https://doi.org/10.1007/s12021-012-9151-4>
- Hannon, B., Mandy, W., & Hull, L. (2023). A comparison of methods for measuring camouflaging in autism. *Autism Research*, 16(1), 12–29. <https://doi.org/10.1002/aur.2850>
- Harris, L., Gilmore, D., Longo, A., & Hand, B. N. (2021). Patterns of US federal autism research funding during 2017–2019. *Autism*, 25(7), 2135–2139. <https://doi.org/10.1177/13623613211003430>
- Harrop, C., Dallman, A. R., & Boyd, B. A. (2021). Big data. Big problem? *Autism Research*, 14(2), 238–239. <https://doi.org/10.1002/aur.2465>
- Hughes, D. J. (2018). Psychometric validity. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing* (pp. 751–779). John Wiley & Sons. <https://doi.org/10.1002/9781118489772.ch24>
- Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*, 54(2), 216–224. <https://doi.org/10.1111/j.1469-7610.2012.02589.x>
- Jebb, A. T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Frontiers in Psychology*, 12, 637547. <https://doi.org/10.3389/fpsyg.2021.637547>
- Keating, C. T. (2021). Participatory autism research: How consultation benefits everyone. *Frontiers in Psychology*, 12, 713982. <https://doi.org/10.3389/fpsyg.2021.713982>
- Kim, S. Y., & Lecavalier, L. (2022). Evaluating the use of self-reported measures in autistic individuals in the context of psychiatric assessment: A systematic review. *Journal of Autism and Developmental Disorders*, 52(10), 4355–4374. <https://doi.org/10.1007/s10803-021-05323-y>
- Kunin, T. (1955). The construction of a new type of attitude measure 1. *Personnel Psychology*, 8(1), 65–77. <https://doi.org/10.1111/j.1744-6570.1955.tb01189.x>
- Lai, M., Hull, L., Mandy, W., Chakrabarti, B., Nordahl, C. W., Lombardo, M. V., Ameis, S. H., Szatmari, P., Baron-Cohen, S., & Happé, F. (2021). Commentary: ‘Camouflaging’ in autistic people—reflection on Fombonne (2020). *Journal of Child Psychology and Psychiatry*, 62(8), 1037–1041. <https://doi.org/10.1111/jcpp.13344>
- Lewis, C. C., Puspitasari, A., Boyd, M. R., Scott, K., Marriott, B. R., Hoffman, M., Navarro, E., & Kassab, H. (2018). Implementing measurement based care in community mental health: A description of tailored and standardized methods. *BMC Research Notes*, 11(1), 1–6. <https://doi.org/10.1186/s13104-018-3193-0>
- Lyall, K., Rando, J., Toroni, B., Ezeh, T., Constantino, J. N., Croen, L. A., Garvin, B., Piselli, K., Connell, J., & Kaat, A. J. (2022). Examining shortened versions of the social responsiveness scale for use in autism spectrum disorder prediction and as a quantitative trait measure: Results from a validation study of 3–5 year old children. *JCPP Advances*, 2, e12106. <https://doi.org/10.1002/jcv2.12106>
- Magiati, I., Lerh, J. W., Hollocks, M. J., Uljarevic, M., Rodgers, J., McConachie, H., Ozsivadjian, A., South, M., Van Hecke, A., Hardan, A., Libove, R., Leekam, S., & Simonoff, E. (2017). The measurement properties of the Spence Children’s Anxiety Scale—parent version in a large international pooled sample of young people with autism spectrum disorder. *Autism Research*, 10(10), 1629–1652. <https://doi.org/10.1002/aur.1809>
- Mazefsky, C. A., Day, T. N., Siegel, M., White, S. W., Yu, L., & Pilkonis, P. A. (2018). Development of the emotion dysregulation inventory: A PROMIS® ing method for creating sensitive and unbiased questionnaires for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 48(11), 3736–3746. <https://doi.org/10.1007/s10803-016-2907-1>
- Mazefsky, C. A., Herrington, J., Siegel, M., Scarpa, A., Maddox, B. B., Scahill, L., & White, S. W. (2013). The role of emotion regulation in autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(7), 679–688. <https://doi.org/10.1016/j.jaac.2013.05.006>
- Mazefsky, C. A., Yu, L., & Pilkonis, P. A. (2021). Psychometric properties of the emotion dysregulation inventory in a nationally representative sample of youth. *Journal of Clinical Child & Adolescent Psychology*, 50(5), 596–608. <https://doi.org/10.1080/15374416.2019.1703710>
- Mazefsky, C. A., Yu, L., White, S. W., Siegel, M., & Pilkonis, P. A. (2018). The emotion dysregulation inventory: Psychometric properties and item response theory calibration in an autism spectrum disorder sample. *Autism Research*, 11(6), 928–941. <https://doi.org/10.1002/aur.1947>
- McConachie, H., Wilson, C., Mason, D., Garland, D., Parr, J. R., Rattazzi, A., Rodgers, J., Skevington, S., Uljarevic, M., & Magiati, I. (2020). What is important in measuring quality of life? Reflections by autistic adults in four countries. *Autism in Adulthood*, 2(1), 4–12. <https://doi.org/10.1089/aut.2019.0008>
- McDermott, C. R., Farmer, C., Gotham, K. O., & Bal, V. H. (2020). Measurement of subcategories of repetitive behaviors in autistic adolescents and adults. *Autism in Adulthood*, 2(1), 48–60. <https://doi.org/10.1089/aut.2019.0056>
- McFayden, T. C., Gatto, A. J., Dahiya, A. V., Antezana, L., Miyazaki, Y., & Cooper, L. D. (2021). Integrating measurement-based care into treatment for autism spectrum disorder: Insights from a community clinic. *Journal of Autism and Developmental Disorders*, 51(10), 3651–3661. <https://doi.org/10.1007/s10803-020-04824-6>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, 5, e16800. <https://doi.org/10.7554/eLife.16800>
- McNeish, D. (2018). Thanks coefficient alpha, we’ll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Mokkink, L. B., Boers, M., van der Vleuten, C. P. M., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2020). COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: A Delphi study. *BMC Medical Research Methodology*, 20(1), 293. <https://doi.org/10.1186/s12874-020-01179-5>
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported

- outcome measures. *Quality of Life Research*, 27(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Mokkink, L. B., Prinsen, C. A. C., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2016). The COnsensus-based standards for the selection of health measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy*, 20(2), 105–113. <https://doi.org/10.1590/bjpt-rbf.2014.0143>
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737–745. <https://doi.org/10.1016/j.jclinepi.2010.02.006>
- Newman, J. C., & Feldman, R. (2011). Copyright and open access at the bedside. *New England Journal of Medicine*, 365(26), 2447–2449. <https://doi.org/10.1056/NEJMp1110652>
- Nicolaidis, C., Raymaker, D. M., McDonald, K. E., Lund, E. M., Leotti, S., Kapp, S. K., Katz, M., Beers, L. M., Kripke, C., & Maslak, J. (2020). Creating accessible survey instruments for use with autistic adults and people with intellectual disability: Lessons learned and recommendations. *Autism in Adulthood*, 2(1), 61–76. <https://doi.org/10.1089/aut.2019.0074>
- Nicolaidis, C., Zhen, K. Y., Lee, J., Raymaker, D. M., Kapp, S. K., Croen, L. A., Urbanowicz, A., Maslak, J., & Scharer, M. (2021). Psychometric testing of a set of patient-reported instruments to assess healthcare interventions for autistic adults. *Autism*, 25(3), 786–799. <https://doi.org/10.1177/1362361320967178>
- Nolte, S., Coon, C., Hudgens, S., & Verdum, M. G. E. (2019). Psychometric evaluation of the PROMIS® Depression Item Bank: An illustration of classical test theory methods. *Journal of Patient-Reported Outcomes*, 3(1), 46. <https://doi.org/10.1186/s41687-019-0127-0>
- Pecora, L. A., Hooley, M., Sperry, L., Mesibov, G. B., & Stokes, M. A. (2020). Sexuality and gender issues in individuals with autism spectrum disorder. *Child and Adolescent Psychiatric Clinics*, 29(3), 543–556. <https://doi.org/10.1016/j.chc.2020.02.007>
- Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, 18(1), 25–34. <https://doi.org/10.1016/j.jval.2014.2014.05>
- Poulsen, R., Dwyer, P., Gassner, D., Heyworth, M., & Williams, Z. J. (2023). The INSAR community collaborator request: Using community–academic partnerships to enhance outcomes of participatory autism research. *Autism Research*, 16, 2071–2076. <https://doi.org/10.1002/aur.3027>
- Preacher, K. J., & MacCallum, R. C. (2003). Repairing tom swift's electric factor analysis machine. *Understanding Statistics*, 2(1), 13–43. https://doi.org/10.1207/S15328031US0201_02
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Prinsen, C. A. C., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., Williamson, P. R., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—A practical guideline. *Trials*, 17(1), 449. <https://doi.org/10.1186/s13063-016-1555-2>
- Ratto, A. B., Bascom, J., DaVanport, S., Strang, J. F., Anthony, L. G., Verbalis, A., Pugliese, C., Nadwodny, N., Brown, L. X. Z., & Cruz, M. (2022). Centering the inner experience of autism: Development of the self-assessment of autistic traits. *Autism in Adulthood*, 5, 93–105. <https://doi.org/10.1089/aut.2021.0099>
- Raykov, T., & Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, 79(1), 200–210. <https://doi.org/10.1177/0013164417725127>
- Revelle, W., & Condon, D. M. (2019). Reliability from α to ω : A tutorial. *Psychological Assessment*, 31(12), 1395–1411. <https://doi.org/10.1037/pas0000754>
- Riccio, A., Delos Santos, J., Kapp, S. K., Jordan, A., DeNigris, D., & Gillespie-Lynch, K. (2020). Developing the multidimensional visual scale assessing affect, anxiety, pride, and energy through a research partnership with autistic scholars. *Autism in Adulthood*, 2(1), 87–100. <https://doi.org/10.1089/aut.2019.0067>
- Rødgaard, E., Jensen, K., Miskowiak, K. W., & Mottron, L. (2022). Representativeness of autistic samples in studies recruiting through social media. *Autism Research*, 15(8), 1447–1456.
- Rodgers, J., Farquhar, K., Mason, D., Brice, S., Wigham, S., Ingham, B., Freeston, M., & Parr, J. R. (2020). Development and initial evaluation of the anxiety scale for autism-adults. *Autism in Adulthood*, 2(1), 24–33. <https://doi.org/10.1089/aut.2019.0044>
- Rodgers, J., Wigham, S., McConachie, H., Freeston, M., Honey, E., & Parr, J. R. (2016). Development of the anxiety scale for children with autism spectrum disorder (ASC-ASD). *Autism Research*, 9, 1205–1215. <https://doi.org/10.1002/aur.1603>
- Sasson, N. J., Faso, D. J., Parlier, M., Daniels, J. L., & Piven, J. (2014). When father doesn't know best: Selective disagreement between self-report and informant report of the broad autism phenotype in parents of a child with autism. *Autism Research*, 7(6), 731–739. <https://doi.org/10.1002/aur.1425>
- Sauer, J., Baumgartner, J., Frei, N., & Sonderegger, A. (2020). Pictorial scales in research and practice. *European Psychologist*, 26, 112–130. <https://doi.org/10.1027/1016-9040/a000405>
- Scahill, L., Lecavalier, L., Schultz, R. T., Evans, A. N., Maddox, B., Pritchett, J., Herrington, J., Gillespie, S., Miller, J., Amoss, R. T., Aman, M. G., Bearss, K., Gadow, K., & Edwards, M. C. (2019). Development of the Parent-Rated Anxiety Scale for youth with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58(9), 887–896. <https://doi.org/10.1016/j.jaac.2018.10.016>
- Schiltz, H. K., & Magnus, B. E. (2020). Gender-based differential item functioning on the child behavior checklist in youth on the autism spectrum: A brief report. *Research in Autism Spectrum Disorders*, 79, 101669. <https://doi.org/10.1016/j.rasd.2020.101669>
- Schiltz, H. K., & Magnus, B. E. (2021). Differential item functioning based on autism features, IQ, and age on the screen for child anxiety related disorders (scared) among youth on the autism spectrum. *Autism Research*, 14(6), 1220–1236. <https://doi.org/10.1002/aur.2481>
- Schiltz, H. K., Magnus, B. E., McVey, A. J., Haendel, A. D., Dolan, B. K., Stanley, R. E., Willar, K. A., Pleiss, S. J., Carson, A. M., & Carlson, M. (2019). A psychometric analysis of the social anxiety scale for adolescents among youth with autism spectrum disorder: Caregiver-adolescent agreement, factor structure, and validity. *Assessment*, 28(1), 100–115. <https://doi.org/10.1177/1073191119851563>
- Schiltz, H. K., McVey, A. J., Lord, C. (in press). *Anxiety disorders in autistic people: A narrative review*. Psychiatric Clinics.
- Schwartzman, J. M., Williams, Z. J., Paterson, A. V., Jacobs, A. X., & Corbett, B. A. (2023). Community-guided measurement-based care for autistic youth and adults receiving psychotherapy: A conceptual overview and pilot implementation study of MBC-AUT. *Autism*, 13623613221143587, 1658–1675. <https://doi.org/10.1177/13623613221143587>
- Shogren, K. A., Anderson, M. H., Raley, S. K., & Hagiwara, M. (2021). Exploring the relationship between student and teacher/proxy-responder scores on the Self-Determination Inventory. *Exceptionality*, 29(1), 47–60. <https://doi.org/10.1080/09362835.2020.1729764>
- Smid, S. C., McNeish, D., Miočević, M., & van de Schoot, R. (2020). Bayesian versus frequentist estimation for structural equation models in small sample contexts: A systematic review. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 131–161. <https://doi.org/10.1080/10705511.2019.1577140>

- Stover, A. M., McLeod, L. D., Langer, M. M., Chen, W.-H., & Reeve, B. B. (2019). State of the psychometric methods: Patient-reported outcome measure development and refinement using item response theory. *Journal of Patient-Reported Outcomes*, 3, 1–16. <https://doi.org/10.1186/s41687-019-0130-5>
- Sturm, A., Kuhfeld, M., Kasari, C., & McCracken, J. T. (2017). Development and validation of an item response theory-based Social Responsiveness Scale short form. *Journal of Child Psychology and Psychiatry*, 58(9), 1053–1061. <https://doi.org/10.1111/jcpp.12731>
- Taylor, S. C., Gehringer, B. N., Dow, H. C., Langer, A., Rawot, E., Smernoff, Z., Steeman, S., Almasy, L., Rader, D. J., & Bučan, M. (2022). Contrasting views of autism spectrum traits in adults, especially in self-reports vs. informant-reports for women high in autism spectrum traits. *Journal of Autism and Developmental Disorders*, 1–13. <https://doi.org/10.1007/s10803-022-05822-6>
- ten Hove, D., Jorgensen, T. D., & van der Ark, L. A. (2022). Updated guidelines on selecting an intraclass correlation coefficient for interrater reliability, with applications to incomplete observational designs. *Psychological Methods*. <https://doi.org/10.1037/met0000516>
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, 27(5), 1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>
- Tomaszewski, B., Kraemer, B., Steinbrenner, J. R., Smith DaWalt, L., Hall, L. J., Hume, K., & Odom, S. (2020). Student, educator, and parent perspectives of self-determination in high school students with autism spectrum disorder. *Autism Research*, 13(12), 2164–2176. <https://doi.org/10.1002/aur.2337>
- Ulitzsch, E., Lüdtke, O., & Robitzsch, A. (2023). Alleviating estimation problems in small sample structural equation modeling—A comparison of constrained maximum likelihood, Bayesian estimation, and fixed reliability approaches. *Psychological Methods*, 28(3), 527–557. <https://doi.org/10.1037/met0000435>
- Uljarević, M., Frazier, T. W., Jo, B., Scahill, L., Youngstrom, E. A., Spackman, E., Phillips, J. M., Billingham, W., & Hardan, A. (2022). Dimensional assessment of restricted and repetitive behaviors: Development and preliminary validation of a new measure. *Journal of the American Academy of Child & Adolescent Psychiatry*, 62, 568–581. <https://doi.org/10.1016/j.jaac.2022.07.863>
- Vitoratou, S., Uglik-Marucha, N., Hayes, C., & Pickles, A. (2024). A comprehensive guide for assessing measurement tool quality: The Contemporary Psychometrics (ConPsy) Checklist. <https://doi.org/10.31234/osf.io/t2pbj>
- Wagner, A., Caplan, L., Juliano-Bult, D., & Williams, N. (2020). Improving the rigor of research on autism in adulthood requires valid and reliable measurement tools. In *Autism in adulthood* (Vol. 2, pp. 174–176). Mary Ann Liebert. <https://doi.org/10.1089/aut.2020.0001>
- Williams, Z. J. (2022). Commentary: The construct validity of ‘camouflaging’ in autism: Psychometric considerations and recommendations for future research-reflection on Lai et al. (2020). *Journal of Child Psychology and Psychiatry*, 63(1), 118–121. <https://doi.org/10.1111/jcpp.13468>
- Williams, Z. J., Cascio, C. J., & Woynaroski, T. G. (2023). Measuring subjective quality of life in autistic adults with the PROMIS global-10: Psychometric study and development of an autism-specific scoring method. *Autism*, 27(1), 145–157. <https://doi.org/10.1177/13623613221085364>
- Williams, Z. J., Everaert, J., & Gotham, K. O. (2021). Measuring depression in autistic adults: Psychometric validation of the Beck Depression Inventory–II. *Assessment*, 28(3), 858–876. <https://doi.org/10.1177/1073191120952889>
- Williams, Z. J., Schaaf, R., Ausderau, K. K., Baranek, G. T., Barrett, D. J., Cascio, C. J., Dumont, R. L., Eyoh, E. E., Failla, M. D., Feldman, J. I., Foss-Feig, J. H., Green, H. L., Green, S. A., He, J. L., Kaplan-Kahn, E. A., Keçeli-Kaysılı, B., MacLennan, K., Mailloux, Z., Marco, E. J., ... Woynaroski, T. G. (2023). Examining the latent structure and correlates of sensory reactivity in autism: A multi-site integrative data analysis by the autism sensory research consortium. *Molecular Autism*, 14(1), 31. <https://doi.org/10.1186/s13229-023-00563-4>
- Zickar, M. J. (2020). Measurement development and evaluation. *Annual Review of Organizational Psychology and Organizational Behavior*, 7, 213–232. <https://doi.org/10.1146/annurev-orgpsych-012119-044957>

How to cite this article: Schiltz, H. K., Williams, Z. J., Zheng, S., Kaplan-Kahn, E. A., Morton, H. E., Rosenau, K. A., Nicolaidis, C., Sturm, A., Lord, C., & Autism PROMnet (2024). Measurement matters: A commentary on the state of the science on patient reported outcome measures (PROMs) in autism research. *Autism Research*, 1–12. <https://doi.org/10.1002/aur.3114>