

Studying Development of Psychopathology Across the Lifespan

Isaac T. Petersen, Zachary Demko, and Megan Schumer

Department of Psychological and Brain Sciences, University of Iowa

Author Note

Isaac T. Petersen  <https://orcid.org/0000-0003-3072-6673>

Zachary Demko  <https://orcid.org/0000-0002-5717-6205>

Megan Schumer  <https://orcid.org/0009-0009-5488-6541>

The data and code for the present study are published online:

<https://github.com/DevPsyLab/studying-psychopathology-development-lifespan>

(<https://osf.io/t3zf9>). We have no conflicts of interest to disclose. I.P. was funded by Grant HD098235 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Institutes of Health. We thank Jordan L. Harris and Alexis Hosch for their help identifying measures of externalizing problems and transcribing and coding their content. Author roles were classified using the Contributor Role Taxonomy (CRediT; <https://credit.niso.org>) as follows: Isaac T. Petersen: conceptualization, data curation, analysis, funding, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing, editing; Zachary Demko: conceptualization, data curation, analysis, investigation, methodology, software, editing; Megan Schumer: editing

Correspondence concerning this article should be addressed to Isaac T. Petersen, University of Iowa, G60 Psychological and Brain Sciences Building, Iowa City, IA 52242. Email: isaac-t-petersen@uiowa.edu

Abstract

The field has not adequately studied individuals' development of psychopathology across the lifespan, instead taking a piecewise approach to development, limiting our understanding of lifespan development and the origins of psychopathology. We describe why this is the case—beyond logistical challenges. First, diagnostic and taxonomic systems do not define psychopathology constructs from a developmental perspective. A developmental perspective is crucial because psychopathology constructs often have earlier origins than their full clinical presentation, and forms of psychopathology frequently show changes in behavioral manifestation across development (heterotypic continuity). Second, current assessments do not adequately capture these changes in behavioral manifestation. Third, traditional scoring systems do not allow charting individuals' change over time in the face of changing behavioral manifestation. Fourth, assessments are not well-suited for capturing individual differences, despite the dimensional nature of psychopathology. We propose solutions to these challenges by introducing a developmental scaling framework for psychopathology. Using externalizing psychopathology as an example, we draw on our content analysis of externalizing measures, as well as lessons from the study of other constructs across lengthy developmental periods. Addressing these challenges will allow charting individuals' trajectories of psychopathology across the lifespan and identifying risk and protective factors that influence developmental courses.

Keywords: psychopathology, lifespan development, longitudinal trajectories, heterotypic continuity, externalizing behavior problems

Studying Development of Psychopathology Across the Lifespan

NOTE: this manuscript is in progress of being drafted—it is incomplete.

When entrenched, psychopathology is costly and difficult to treat (Insel, 2008; Wakschlag et al., 2019). Understanding its developmental origins is therefore crucial for preventing later, more severe psychopathology. Developmental science seeks to build a bridge that spans the lifespan—from infancy to older adulthood—and not just transitory periods in people's lives. However, the field has not adequately studied individuals' development of psychopathology across the lifespan; instead, the field has taken a piecewise approach to development that focuses on narrower age ranges. The focus on narrow age ranges restricts our ability to understand development across important transitions, such as puberty, school entry, and entry into parenthood, which in turn limits our understanding of lifespan development and the origins of psychopathology. In this paper, we describe reasons why this is the case—beyond logistical challenges of time and money. And we propose solutions to these challenges by introducing a developmental scaling framework for psychopathology. Using externalizing psychopathology as an example, we draw on our content analysis of more than 270 externalizing measures, as well as lessons from the study of other constructs such as academic and cognitive skills across lengthy developmental periods.

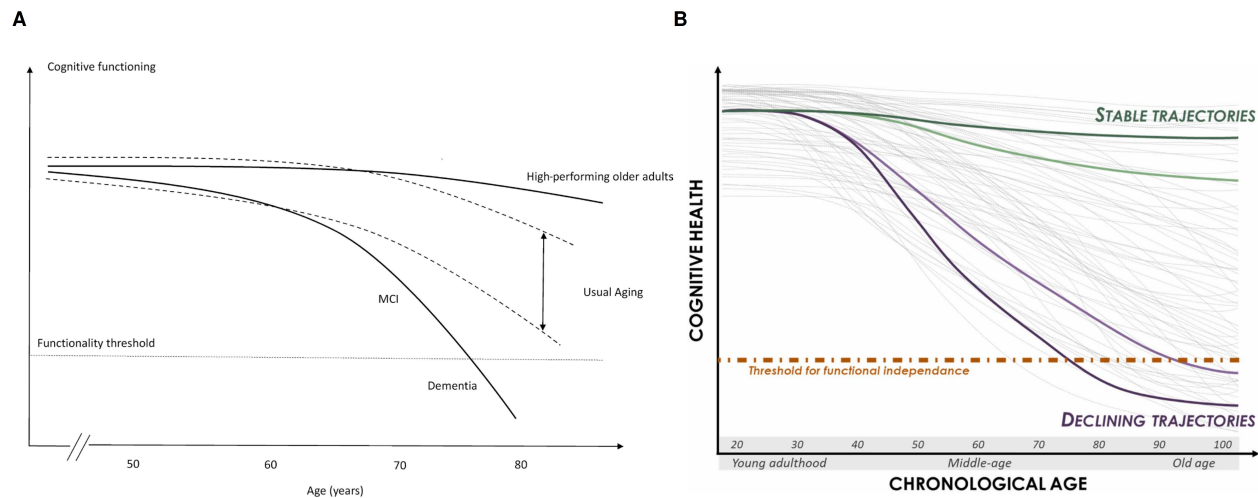
The Goal

A key goal of developmental science is to understand how people develop across the lifespan. A variety of methods, including cross-sectional and longitudinal designs, could be leveraged for this purpose. In a cross-sectional design, people are assessed at one timepoint. In a longitudinal design, the same individuals are followed across time.

Cross-sectional designs provide the ability to relatively cheaply and quickly assess a wide age range, providing a snapshot at a particular point in time. Cross-sectional designs may inform our understanding of general normative patterns of change—such as the typical pattern of cognitive growth and decline across the lifespan. Thus, cross-sectional designs provide an important starting point.

However, cross-sectional designs have key limitations. First, cross-sectional designs do not allow examining individuals' change over time. Thus, the trajectory of individual differences remains elusive, and observed age-related differences may reflect processes apart from developmental change, such as cohort effects. Second, cross-sectional designs may obscure diverse patterns of change. Individuals show considerable heterogeneity in level and change across time in many constructs. As an example, despite the normative pattern of steep cognitive decline in episodic memory with aging, there are a subset of individuals who do not show this same pattern. For instance, superagers are individuals who are older adults (age 80+) who have episodic memory at or above the normative level of middle-aged adults (50–65 years old) and who may be resilient to the typical aging process ([Godoy et al., 2021](#)). Theoretical depictions of the heterogeneity of cognitive decline are depicted in Figure 1.

Figure 1
Different Cognitive Trajectories of Aging



Note. Panel A: Hypothetical models for different cognitive trajectories of aging. MCI = mild cognitive impairment. (Figure in Panel A reprinted from Borelli et al. (2018), Figure 1, p. 222. Borelli, W. V., Carmona, K. C., Studart-Neto, A., Nitrini, R., Caramelli, P., & Costa, J. C. d. (2018). Operationalized definition of older adults with high cognitive performance. *Dementia & Neuropsychologia*, 12(3), 221–227. <https://doi.org/10.1590/1980-57642018dn12-030001>). **Panel B:** Theoretical illustration of inter-individual variability in cognitive aging trajectories. Each gray line represents one individual hypothetical cognitive health trajectory. The orange line marks the threshold for functional independence; once an individual's cognitive health drops below this level, autonomous daily living is compromised. (Figure in Panel B reprinted from Abrous et al. (2026), Figure 1, p. 2. Abrous, D. N., Blin, N., Boraxbekk, C.-J., Catheline, G., Fitzsimons, C. P., Hilscher, M., Lemoine, M., Lopes, L. V., Maass, A., Nilsson, M., Nyberg, L., Rasmussen, L. J., Remondes, M., Sauvage, M., Schreiber, S., & Wolbers, T. (2026). Hallmarks of healthy cognitive aging: Inter-individual differences in aging trajectories. *Ageing Research Reviews*, 119, 103102. <https://doi.org/10.1016/j.arr.2026.103102>).

Source: Analyses

Heterogeneity also applies to psychopathology trajectories including externalizing and criminal behavior. The age-crime curve depicts the well-known strong association between age and engagement in crime, with crime rates peaking markedly (in many Western countries) in adolescence around ages 15–19. An example of the age-crime curve is depicted in Figure 2.

However, Moffitt's (1993) developmental taxonomy of externalizing (“antisocial”) behavior posits that there are heterogeneous subgroups such that one subgroup exhibits low levels of externalizing behavior across the lifespan, an “adolescence limited” subgroup exhibits high

levels of externalizing behavior during adolescence but low levels during childhood and adulthood, and a “life-course persistent” subgroup that exhibits high levels of externalizing behavior across the lifespan; in addition, more recent work has identified a “childhood limited” subgroup that exhibits high levels of externalizing behavior during childhood but low levels after childhood (Fairchild et al., 2013; Odgers et al., 2008). The developmental taxonomy is depicted in Figure 2.

Moffitt’s general hypothesis is that life-course persistent externalizing behavior results from early disruptions in brain development either in utero (resulting from, for example, maternal substance use, poor prenatal nutrition, genetics, or exposure to toxins), during delivery (e.g., delivery complications), or after birth (e.g., child abuse or neglect, or exposure to toxins) that lead to impaired cognitive abilities and difficult temperament, in perhaps conjunction with coercive parent–child interactions or other adverse, criminogenic environments. Adolescence-limited externalizing behavior, by contrast, is thought to reflect a more normative—even adaptive—pattern of social behavior that is driven by exposure to deviant peers and a maturity gap. Childhood-limited externalizing behavior, though not hypothesized by the original taxonomy, may in some cases reflect recovery and in other cases reflect a shift from externalizing problems into problems in other domains (Odgers et al., 2008). It has been hypothesized that both childhood-limited and life-course persistent externalizing behavior may be characterized by high levels of individual-level risk (e.g., genetic and neuropsychological risk), whereas especially life-course persistent externalizing behavior may be characterized by greater environmental risk (e.g., adversity and trauma; Fairchild et al., 2013). A cross-sectional study that does not capture the subgroups’ different developmental histories would not be well-positioned to adjudicate whether a particular individual is part of the adolescence-limited versus life-course persistent developmental course (Moffitt, 1993).

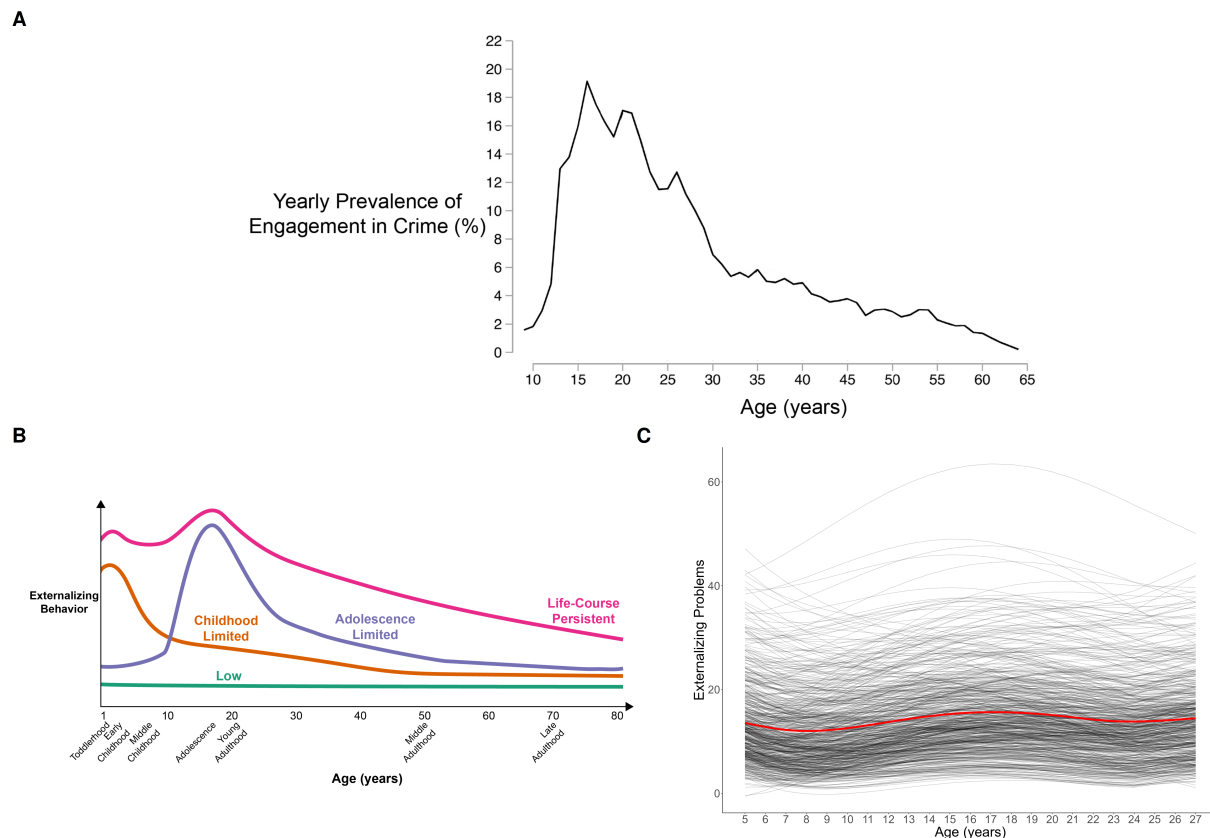
The age-crime curve that is based on official police records has key problems (Moffitt, 1993). First, it misses the developmental course of externalizing behavior that leads up to the commission of crime. Second, it is based on only those (criminalized) behaviors that are known to official arrest records and does not capture other behaviors the person engaged in that are known

to the individual or collateral report by caregivers, teachers, siblings, or peers. Longitudinal work examining informant report of children's behavior allows estimating the prototypical trajectory of externalizing problems at ages younger than those encompassed by the age-crime curve, thus providing important developmental context to the origins of externalizing behavior. For instance, in a longitudinal study examining the development of externalizing behavior from childhood to adulthood that incorporated ratings of the child's externalizing behavior from mothers, fathers, teachers, peers, and self-report (at relevant ages), Petersen et al. (2015) found that externalizing problems normatively tended to decrease from early childhood to preadolescence, increased during adolescence (peaking around age 17), and decreased from late adolescence to adulthood.

Since Moffitt (1993) proposed the developmental taxonomy of externalizing behavior, work has examined individual differences in externalizing trajectories to determine whether they are best represented by categories—i.e., taxa such as adolescence-limited, life-course persistent, etc.—or dimensionally. Findings have consistently shown that variation in externalizing trajectories is better represented dimensionally than with categories (Fairchild et al., 2013; Walters, 2011, 2012, 2015; Walters & Ruscio, 2013), suggesting that differences between people in their developmental course of externalizing behavior appear to reflect differences in degree rather than kind (i.e., quantitative differences rather than qualitative differences). For example, the trajectories in Petersen et al. (2015) were characterized by marked individual differences; some individuals stayed low, some increased, some decreased, some increased then decreased, and seemingly everything in between. The heterogeneous trajectories of externalizing problems are depicted in Figure 2.

In sum, if we rely solely on normative age-related patterns of cognitive functioning or psychopathology from cross-sectional research, we would miss the immense differences in people's developmental courses. And these individual differences are crucial to identify the risk and protective mechanisms that explain these differing developmental courses.

Figure 2
Externalizing Behavior Across Development



Note. Panel A: The age-crime participation curve from age 9 through 64 in cohort members registered for antisocial and/or criminal behavior in a birth cohort recruited in Stockholm, Sweden ($N = 4,825$), based on official crime records. The steep peak in crime rate in the adolescent years is similar to the age-crime curve in the United States; however, the steepness of the age-crime curve differs within country across birth cohorts (Steffensmeier et al., 2025). (Figure adapted from Sivertsson et al. (2024), Figure 1, p. 7. Sivertsson, F., Carlsson, C., Almquist, Y. B., & Brännström, L. (2024). Offending trajectories from childhood to retirement age: Findings from the Stockholm birth cohort study. *Journal of Criminal Justice*, 91, 102155. <https://doi.org/10.1016/j.jcrimjus.2024.102155>).

Panel B: A stylized depiction of Moffitt's (1993) developmental taxonomy of externalizing behavior, adapted to include a childhood limited subgroup. **Panel C:** Individuals' ($N = 585$) longitudinal trajectories of externalizing problems from childhood to adulthood based on annual assessments of mother-, father-, teacher-, peer-, and self-report (at relevant ages). Average trajectory in red. (Figure adapted from Petersen et al. (2015), Figure 2, Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. *Development and Psychopathology*, 27(3), 791–818. <https://doi.org/10.1017/S0954579414000789>).

Source: Analyses

Third, in a cross-sectional study, any apparent age-related differences in a construct could

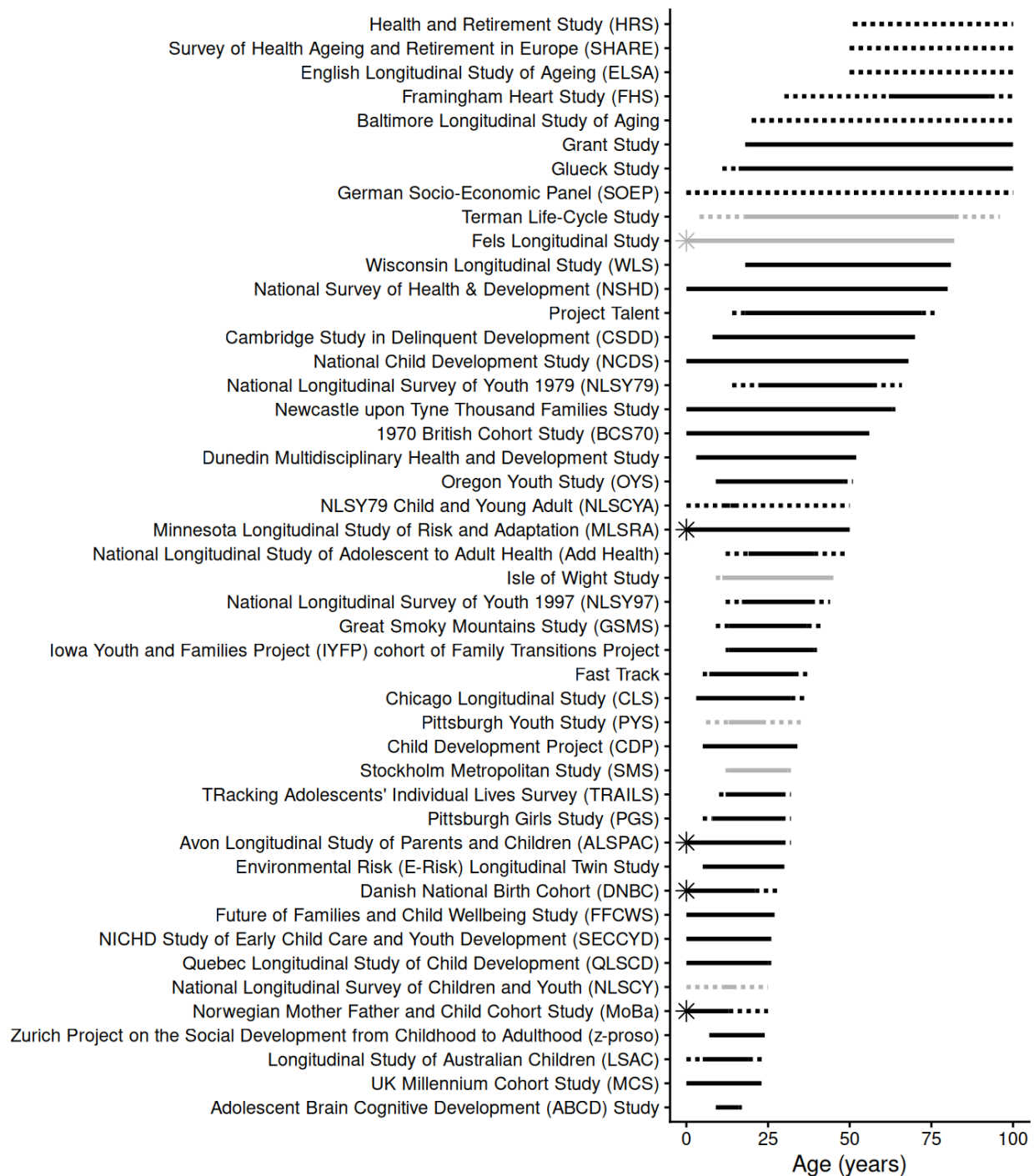
reflect cohort differences instead of developmental changes. For example, average differences in stress between 5-year-olds and 10-year-olds could reflect the differing experiences that the 10-year-olds shared, such as going to school during the Coronavirus pandemic. Fourth, findings from cross-sectional studies can differ from findings that are based on examining the same people over time—a violation of the convergence assumption. In cross-sectional studies, the convergence assumption is the assumption that cross-sectional age differences and longitudinal age changes converge onto a common trajectory. An example where the convergence assumption has been shown to be violated is in studies of cognitive decline. Cross-sectional studies overestimate age-related cognitive declines compared to following the same people over time in a longitudinal study, which may reflect cohort effects such as the Flynn effect (where population intelligence scores at a given age tend to increase across generations; [Ackerman, 2013](#)).

Because longitudinal designs follow the same person(s) over time, such a design allows examining individuals' change across time. However, as we describe [later](#), longitudinal designs also have potential confounds of change. Moreover, compared to cross-sectional designs, longitudinal designs take more time and tend to be more costly and require more personnel. Thus, longitudinal studies that span the lifespan are not common. However, even cross-sectional work spanning the lifespan is not often conducted.

The Problem/Obstacle

The focal problem examined in this paper is that, despite calls to do so ([De Los Reyes, 2026](#)), the field has not adequately studied individuals' development of psychopathology across the lifespan, instead taking a piecewise approach to development. The piecewise approach to development limits our understanding of lifespan development and the origins of psychopathology.

A modest number of longitudinal studies have followed people over lengthy developmental periods. We overview some of the seminal longitudinal studies that cover a lengthy span of development. A depiction of the ages spanned by various longitudinal studies is in [Figure 3](#).

Figure 3*(Approximate) Ages Spanned by Various Lengthy Longitudinal Studies*

Note. The list of longitudinal studies is as of this writing and is not exhaustive. Studies that are no longer active are shown in gray. Solid lines reflect fully longitudinal spans (i.e., the same individuals were followed over time). Dotted lines reflect ages assessed cross-sectionally—even if some participants may have been followed longitudinally. For instance, if a study recruited 5–10-year-old children and followed each child for 20 years, the study would have dotted lines from ages 5–10 and from ages 25–30. Studies shown entirely with dotted lines assessed their full age range cross-sectionally but still included longitudinal follow-up of some participants. A star at age 0 represents that participants were enrolled before birth and the study included prenatal assessments. Ages spanned reflect ages when assessments were collected directly from participants; ages spanned do not include links to registry data (e.g., birth records, offending records). Age was truncated at 100;

Source: [Analyses](#)

Seminal Longitudinal Studies

Fels Longitudinal Study

The Fels Longitudinal Study ([Roche, 1992](#)) was a cohort-sequential study lasting from 1929 to 2018 that recruited 10–20 pregnant mothers each year from the Yellow Springs, Ohio area, and followed their newborns over time, eventually accruing 2,567 child participants and following them into older adulthood (e.g., age 82; [Chumlea et al., 2012](#)). The study collected data on participants' body composition, including height and weight, which had key impacts and formed the basis of pediatric growth charts. In addition, the study collected information on the participants' behavior—and other constructs—in childhood (based on observation) and adulthood (based on interview). These rich data allowed researchers to examine the extent of personality continuity and change from childhood to adulthood, and the extent to which mothers may influence the personality development of the child ([Moss & Kagan, 1964](#)). On the basis of predictive associations from childhood to adulthood in the Fels Longitudinal Study, Kagan and Moss ([1962](#)) noted that girls who had frequent tantrums at 6 to 10 years of age tended to become women who were more motivated in school, less dependent on others, and more masculine in their interests than women who had fewer tantrums as children. They interpreted the different behaviors at different ages as stemming from the same psychological process: a tendency to avoid adopting “female sex-role standards” (p. 200). Kagan described such findings, when psychological processes remain the same but the form of behavior changes, as examples of heterotypic continuity ([Kagan, 1969, 1971, 1980](#)).

Dunedin Multidisciplinary Health and Development Study

The Dunedin Multidisciplinary Health and Development Study is an ongoing birth cohort study that started in 1975 and recruited 1,037 3-year-old children in Dunedin, New Zealand (who were born in 1972–1973; [Poulton et al., 2015](#)). It has followed participants to age 52 and includes data on a wide range of health and behavior constructs. The study has yielded important insights including that children's self-control is more important than socioeconomic status or IQ in

predicting their health, wealth, and crime in adulthood ([Moffitt et al., 2011](#)), a general psychopathology factor (p factor) may account for the strong covariance among psychological disorders ([Caspi et al., 2014](#)), that mental disorders have earlier developmental origins ([Poulton et al., 2015](#)), a person's mental disorder tends to shift from one disorder to other disorders over the life course ([Caspi et al., 2020](#)), and cross-sectional (i.e., retrospective) studies grossly underestimate lifetime prevalence of psychological disorders (compared to prospective longitudinal studies; [Schaefer et al., 2017](#)).

Harvard Study of Adult Development

The Harvard Study of Adult Development consists of the Grant Study ([Vaillant, 2012](#)) and the Glueck Study ([Glueck & Glueck, 1950](#)). The Grant Study started in 1938 and recruited 268 White men from the Harvard classes of 1939–1944 at ages 18–19, and has followed them into their mid-90s. The Glueck Study started in 1940 and recruited 456 11–16-year-old boys who were living in poor, high-crime inner-city neighborhoods in Boston, Massachusetts, and has followed them into their 80s. One of the key findings from the Harvard Study of Adult Development, summarized in a Harvard Gazette article ([Mineo, 2017](#)), is the importance of close relationships for healthy aging: “Close relationships, more than money or fame, are what keep people happy throughout their lives, the study revealed. Those ties protect people from life's discontents, help to delay mental and physical decline, and are better predictors of long and happy lives than social class, IQ, or even genes...people's level of satisfaction with their relationships at age 50 was a better predictor of physical health than their cholesterol levels were.”

Child Development Project

The Child Development Project ([Dodge et al., 1990](#)) is a study that recruited 4-year-old children the year before kindergarten in 1987 and 1988 in Nashville, Tennessee, Knoxville, Tennessee, and Bloomington, Indiana. Participants included 585 children, and they have been followed to age 34. One of the key findings from the Child Development Project is that a hostile attribution bias, an aspect of social information processing that involves the tendency to interpret ambiguous social cues as hostile, predicts later aggression and that part of the association is

mediated by peer rejection ([Lansford et al., 2010](#)).

Longitudinal Studies of Psychopathology

Longitudinal studies such as these have had immense impacts. These studies have yielded many inferences that would not have been possible without longitudinal designs. Some long-term longitudinal studies have even followed children of those children who were initially recruited, and are thus uniquely positioned to study questions of intergenerational transmission (e.g., [Kim et al., 2009](#)). However, relatively few longitudinal studies attempt to assess people's *change* (in level) in a given psychopathology construct over a lengthy period. Here, we focus on studies that have examined trajectories of externalizing behavior.¹

A depiction of the ages spanned by various longitudinal studies is in Figure 4. In the Child Development Project, Petersen et al. ([2015](#)) examined trajectories of externalizing behavior from ages 5–27 with 20 measurement occasions. In the Dunedin Multidisciplinary Health and Development Study, Odgers et al. ([2008](#)) examined trajectories of antisocial conduct problems from ages 7–26 with 8 measurement occasions. In the Mannheim Study of Children at Risk, Sacu et al. ([in press](#)) examined trajectories of externalizing behavior from ages 8–33 with 8 measurement occasions. A study using six longitudinal datasets examined trajectories of disruptive behaviors and delinquency, with the longest trajectory spanning ages 6–15 with 7 measurement occasions ([Broidy et al., 2003](#)). In addition, a study examined trajectories of externalizing problems from ages 4–17 with 3 measurement occasions in predicting externalizing problems at age 27 ([Korhonen et al., 2018](#)). Other studies have examined trajectories of externalizing problems from ages 5–17 with 5 measurement occasions ([Bista et al., 2025](#)), from ages 6–18 with 11 measurement occasions ([Shi et al., 2020](#)), from ages 1.5–14.5 with 6 measurement occasions ([Kjeldsen et al., 2016](#)), from ages 4–16 with 7 measurement occasions (Avon Longitudinal Study of Parents and Children [ALSPAC]; [Speyer et al., 2022](#)), from ages 12–23 with 3 measurement occasions ([Stringer et al., 2020](#)), from ages 2–15 with 11 measurement

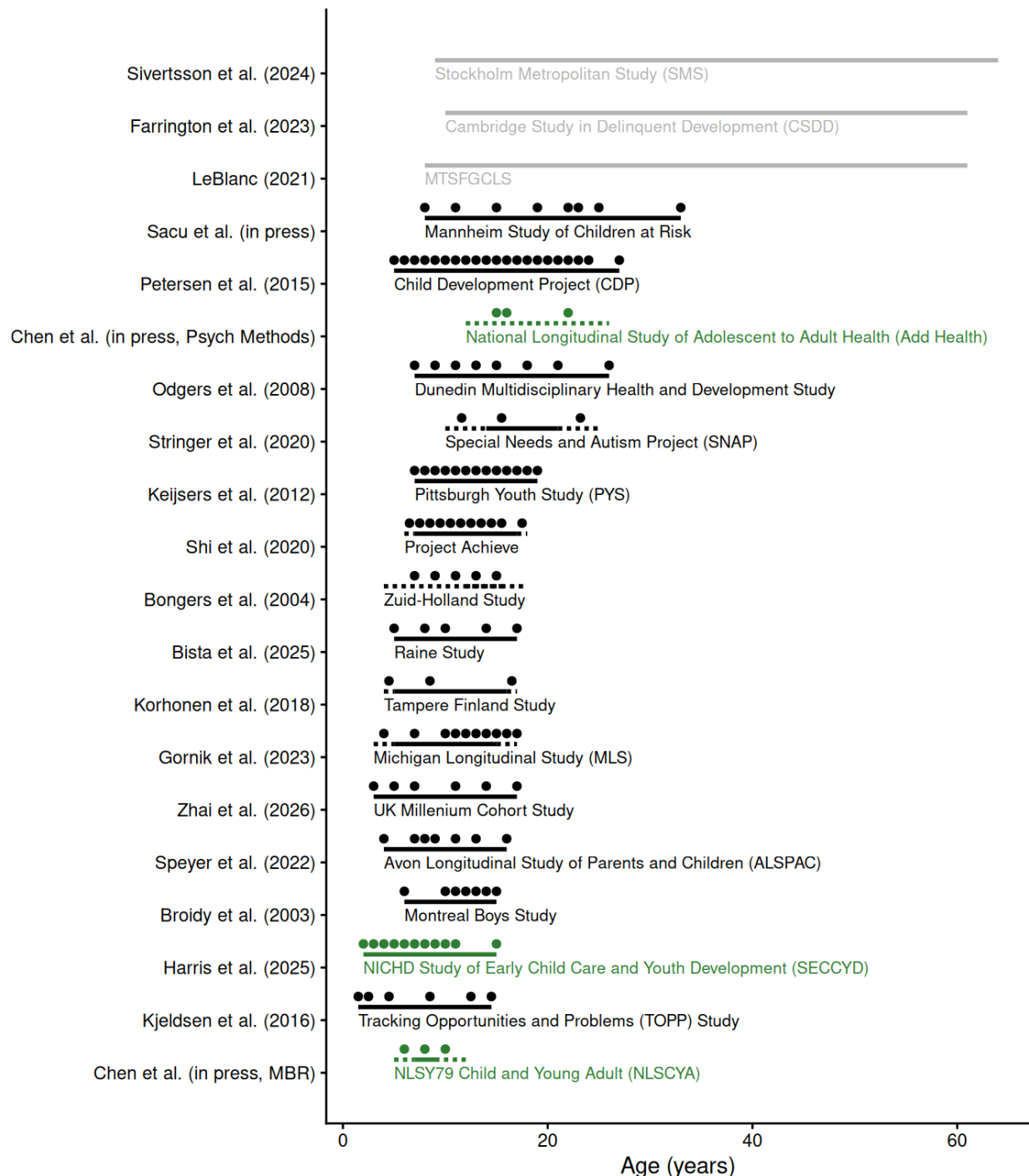
¹ We do not consider autoregressive/cross-lagged models because such models do not examine change in level of externalizing behavior.

occasions (NICHD Study of Early Child Care and Youth Development [SECCYD]; [Harris et al., 2025](#)), and from ages 3–17 with 10 measurement occasions ([Gornik et al., 2023](#)). Bongers et al. (2004) examined trajectories of various externalizing problems from ages 4–18 spanning 13 different birth cohorts with up to 5 measurement occasions per cohort, such that any individual was followed for up to 8 years ([Bongers et al., 2004](#)).

Studies have also examined trajectories of delinquency from ages 7–19 with 13 measurement occasions (Pittsburgh Youth Study; [Keijsers et al., 2012](#)). Demonstrating an analytic method (moderated nonlinear factor analysis) to link scores across measures onto the same scale, Chen and Bauer examined trajectories of delinquency from ages 12–26 with 3 measurement occasions in the National Longitudinal Study of Adolescent to Adult Health (Add Health; [in press b](#)), and examined trajectories of externalizing behavior from ages 5–12 with 3 measurement occasions in the National Longitudinal Survey of Youth 1979 Child and Young Adult (NLSCYA; [in press a](#)). In addition, studies have examined trajectories of emotional and behavioral dysregulation (encompassing externalizing problems and emotional symptoms) from ages 3–17 with 6 measurement occasions (UK Millennium Cohort Study; [Zhai et al., 2026](#)). Studies have also leveraged registry data to examine trajectories of offending from ages 8–61 ([Le Blanc, 2021](#)), from ages 10–61 (Cambridge Study in Delinquent Development; [Farrington et al., 2023](#)), and from ages 9–64 (Stockholm Metropolitan Study; [Sivertsson et al., 2024](#)).

Figure 4

(Approximate) Ages Spanned by Externalizing Trajectories in Various Lengthy Longitudinal Studies Examining Individuals' Trajectories of Externalizing Behavior



Note. The list of longitudinal studies is not exhaustive. Studies that are based on registry data are shown in gray. Studies that used developmental scaling to harmonize scores across ages (i.e., to link scores across ages onto the same scale) are in green. Measurement occasions are depicted with solid circles. Note that, in some studies, measurement occasions only apply to a subset of participants in that study—in some cases, no participant was assessed at all measurement occasions. Solid lines reflect fully longitudinal spans (i.e., the same individuals were followed over time). Dotted lines reflect ages assessed cross-sectionally—even if some participants may have been followed longitudinally. For instance, if a study recruited 5–10-year-old children and followed each child for 20 years, the study would have dotted lines from ages 5–10 and from ages 25–30. Studies

Source: [Analyses](#)

However, as we will describe, these lengthy longitudinal studies of psychopathology (e.g., [Odgers et al., 2008](#))—including some of our own (e.g., [Petersen et al., 2015](#))—have important limitations in (a) how they assessed psychopathology constructs across development and (b) how they estimated people’s scores on those constructs over time, which prevent strong inferences regarding individuals’ change over time.

Reasons for the Problem

Beyond logistical challenges of time and money, there are several key reasons why the field has failed to study people’s development of psychopathology longitudinally across the lifespan. First, diagnostic and taxonomic systems do not define psychopathology constructs from a developmental perspective. A developmental perspective is crucial because psychopathology constructs often have earlier origins than their full clinical presentation, and forms of psychopathology frequently show changes in behavioral manifestation across development (heterotypic continuity). Second, current assessments do not adequately capture these changes in behavioral manifestation. Third, traditional scoring systems do not allow charting individuals’ change over time in the face of changing behavioral manifestation. Fourth, current assessments are not well-suited for capturing the full range of individual differences, despite the dimensional nature of psychopathology. Fifth, because of these limitations, current assessments are not well-suited for capturing individuals’ change over time.

Below, we describe these reasons in detail. To ground the discussion, we provide examples relating to externalizing psychopathology; however, these issues are not specific to externalizing behavior.

Problems of Constructs

The first reason the field has failed to adequately study individuals’ development of psychopathology across the lifespan represents a theory problem—namely, problems with constructs, and the codifying of those constructs in diagnostic and taxonomic systems. Our theories do not fully specify how constructs manifest behaviorally across the lifespan. For instance,

theories do not specify the features (e.g., cognitive, affective, or biological processes) or behaviors that characterize a given construct in each developmental period. In addition, theories do not specify how those features or behaviors change across development with respect to how strongly they reflect the construct—as indicated by changes in factor loadings (in factor analysis) or discrimination (in item response theory)—or how they change in their level on the construct—as indicated by changes in intercepts (factor analysis) or difficulty/severity (item response theory).

These gaps are important for several reasons. First, psychopathology does not come from nowhere; psychopathology constructs often have earlier origins than their full clinical presentation. For instance, children who go on to develop later externalizing problems and conduct problems tend to show earlier temperamental differences such as poorer self-regulation, poorer emotion regulation, low fearfulness, and a more difficult temperament (Bates et al., 2014; Frick & Morris, 2004). In many cases, psychopathology may arise from the interaction of genetic and environmental factors, and their effects on biology, cognition, emotion, and behavior. Such effects may accumulate and transpire over years. Second, forms of psychopathology frequently show changes in behavioral manifestation across development—a phenomenon called heterotypic continuity (Caspi & Shiner, 2006; Cicchetti & Rogosch, 2002; Petersen et al., 2020; Petersen, 2024).

Considerable theoretical work attempts to characterize constructs across development. For instance, Patterson (1993) characterized externalizing (“antisocial”) behavior as a chimera. A chimera is a mythical creature with the body of a goat that, with development, grows the head of a lion and then the tail of a snake. Patterson (1993) and Moffitt (1993) argued that externalizing behavior changes in manifestation across development such that underlying antisociality is maintained while more mature features are added across development. In particular, Patterson (1993) noted that some externalizing behaviors wane with age, such as temper tantrums, whereas other externalizing behaviors emerge with age, including the ability to inflict serious damage or harm with violence, burglary, and shoplifting, in addition to problems such as academic failure, peer rejection, substance problems, and arrest. Moffitt (1993, p. 679) noted that externalizing

behaviors may manifest as “biting and hitting at age 4, shoplifting and truancy at age 10, selling drugs and stealing cars at age 16, robbery and rape at age 22, and fraud and child abuse at age 30”, and that the prognosis for individuals who engage in externalizing behavior throughout the life-course includes “drug and alcohol addiction; unsatisfactory employment; unpaid debts; homelessness; drunk driving; violent assault; multiple and unstable relationships; spouse battery; abandoned, neglected, or abused children; and psychiatric illness”.

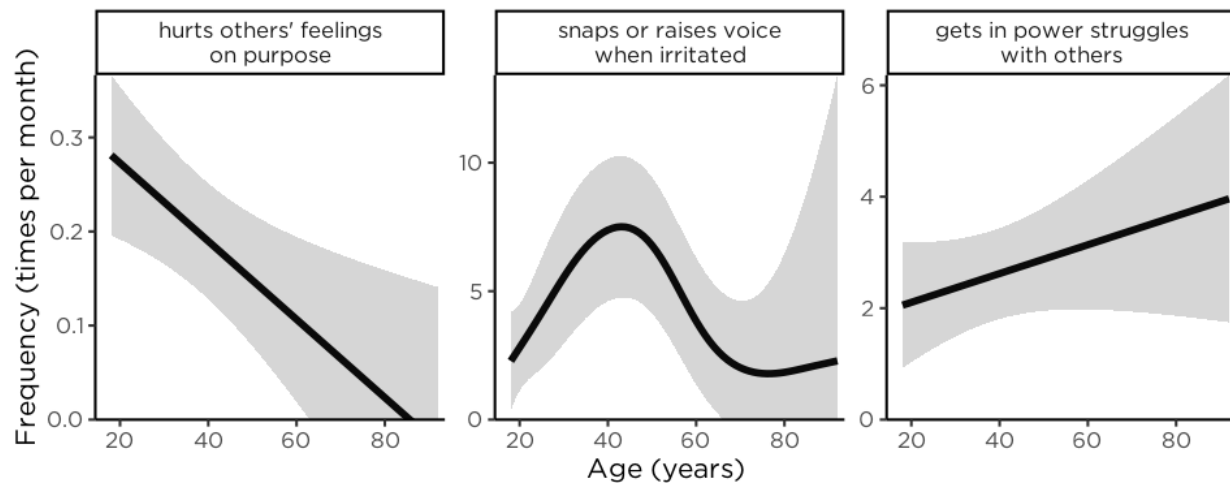
In general, externalizing behaviors are often expressed as overt acts in early childhood, such as physical aggression and temper tantrums; whereas externalizing behaviors are more often expressed later in development as covert and indirect or relational forms of aggression, rule breaking, and substance use (Miller et al., 2009; Petersen, Demko, Lee, et al., 2026). These age-differing behaviors show strong stability of individual differences, indicating a consistent underlying construct (Moffitt, 1993; Patterson, 1993). As depicted in Figure 5, different externalizing behaviors are most common at different developmental periods—even into adulthood, consistent with the notion that externalizing behavior demonstrates heterotypic continuity. Consistent with the Developmental Issues Framework (Sroufe, 2016), the changing expression of externalizing behavior likely reflects a combination of time-varying genetic and environmental factors, such as school entry transition, in combination with varying developmental tasks, different opportunities, and greater experience-dependent capacity (Petersen et al., 2020). These factors likely interact to produce individual differences in heterotypic continuity of relevance to developmental psychopathology, for instance, earlier emergence of alcohol consumption (Aiken et al., 2018).

In addition to externalizing problems, both internalizing (Petersen et al., 2018; Tyrell et al., 2019; Weems, 2008; Weiss & Garber, 2003) and thought-disordered (Rutter et al., 2006) problems are thought to change in manifestation across development.

However, more work is needed to developmentally parameterize constructs—to define constructs developmentally in terms of their features. Recent empirical work suggests that it is uncommon for symptoms to change from one symptom to another for a given person at the

Figure 5

Frequency (Per Month) of Engaging in Various Externalizing Behaviors by Age (N = 1,237)



Note. Items are from the Externalizing Spectrum Inventory (ESI), but unlike the True/False ESI, participants rated items on an absolute frequency scale (times per day/week/month/year). The gray band is the 95% confidence interval. The figure demonstrates that different externalizing behaviors are most common at different developmental periods, consistent with heterotypic continuity. As examples, hurting others' feelings on purpose (left panel) was most common in early adulthood; snapping or raising one's voice (middle panel) was most common in middle adulthood; getting in power struggles with others (right panel) was most common in older adulthood. (Figure reprinted from Petersen, Demko, Doeblen, et al. (2026), Figure 1, p. 301. Petersen, I. T., Demko, Z., Doeblen, P., Sabel, L., Oleson, J. J., & Krueger, R. F. (2026). How often is "often"? Improving assessment of the externalizing spectrum using absolute frequency. *Psychological Assessment*, 38(4), 295–306. <https://doi.org/10.1037/pas0001441>).

subsequent measurement occasion (Applegate & Lahey, 2026). Nevertheless, the authors found that adding new symptoms and subtracting symptoms—other ways that a construct can change in manifestation—were both relatively common. Moreover, just because a symptom persists does not mean that the behavioral manifestation of that symptom stays the same for that individual. For instance, even if a symptom (e.g., “fighting”) persists for an individual, their manifestation of that symptom could differ in form, function, or mechanism (Petersen, 2024). Indeed, fighting is considered to differ in meaning from childhood to adulthood (Patterson, 1993; Tyrell et al., 2019). Thus, as noted by Applegate and Lahey (2026, p. 10), “future causal theories will need to rest on a more detailed and nuanced description of persistence and change at the level of specific psychological problems.”

Moreover, where such theories do indicate changes in behavioral manifestation across development, diagnostic and taxonomic systems do not incorporate these developmental changes.

Lack of a Developmental Perspective

Diagnostic and taxonomic systems do not define psychopathology constructs from a developmental perspective. The diagnostic systems, including the Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR; [American Psychiatric Association, 2022](#)) and the International Classification of Diseases (ICD-11; [World Health Organization, 2022](#)), specify various externalizing-related disorders, including attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD), and antisocial personality disorder (ASPD). Among externalizing-related disorders—with few exceptions²—the diagnostic criteria do not differ across development, even though diagnostic manuals acknowledge that developmental factors impact symptom presentation³. In addition to diagnostic systems, emerging

² In the DSM-5-TR ([American Psychiatric Association, 2022](#)), the minimum frequency for considering a symptom of ODD to be present is “most days” (for a period of six months) for children younger than 5 years, whereas it is “at least once per week” (for at least six months) for children 5 years or older. For the diagnosis of ADHD, six or more symptoms are required for children and younger adolescents (up to age 16), whereas five or more symptoms are required for older adolescents and adults (aged 17 or over). However, the symptoms for defining ODD and ADHD do not differ across development.

³ In the DSM-5-TR ([American Psychiatric Association, 2022](#)), some disorders have a “Development and Course” subsection that describes when disorders tend to emerge and their rates across ages, in addition to describing the types of symptoms that may be more or less common at various points in development. However, the symptoms that define the disorders do not differ across development. For instance, in the section on major depressive disorder, it is noted that “hypersomnia and hyperphagia are more likely in younger individuals, and melancholic symptoms, particularly psychomotor disturbances, are more common in older individuals. Depressions with earlier ages at onset are more familial and more likely to involve personality disturbances.” (p. 189). In the section on ADHD, it is noted that, “In preschool, the main manifestation is hyperactivity. Inattention becomes more prominent during elementary school. During adolescence, signs of hyperactivity (e.g., running and climbing) are less common and may be confined to fidgetiness or an inner feeling of jitteriness, restlessness, or impatience. In adulthood, along with inattention and restlessness, impulsivity may remain problematic even when hyperactivity has diminished.” (p. 71). In the section on CD, it is noted that “Symptoms of the disorder vary with age as the individual develops increased physical strength, cognitive abilities, and sexual maturity. Symptom behaviors that emerge first tend to be less serious (e.g., lying, shoplifting), whereas conduct problems that emerge last tend to be more severe (e.g., rape, theft while confronting a victim)... When individuals with conduct disorder reach adulthood, symptoms of aggression, property destruction, deceitfulness, and rule violation, including violence against co-workers, partners, and children, may be exhibited in the workplace and the home” (p. 534). Moreover, some disorders in the DSM-5-TR note that, to be considered a symptom, the frequency, intensity, or persistence of the behavior must be inconsistent with the level expected for their developmental level. For instance, in the diagnosis of ADHD, it is noted that “the symptoms have persisted for at least 6 months to a degree that is inconsistent with developmental level” (p. 68). In the diagnosis of ODD, it is noted that “other factors [besides a minimum frequency] should also be considered, such as whether the frequency and intensity of the behaviors are outside a range that is normative for the individual’s developmental level, gender, and culture.”

taxonomies and frameworks of psychopathology—including the hierarchical taxonomy of psychopathology (HiTOP) and research domain criteria (RDoC)—do not incorporate a developmental perspective ([Tackett & Hallquist, 2022](#)).

Problems of Existing Assessments and Traditional Scoring Systems

Other key reasons the field has failed to adequately study individuals' development of psychopathology across the lifespan represent methodological problems—namely, problems with existing assessments and how those assessments are used, based on traditional scoring systems. There is a mismatch between theory and method. In particular, there is a mismatch between theory of constructs—including how a construct manifests behaviorally—and the methods used to assess the construct and to evaluate people's level and change in that construct.

Do Not Adequately Capture Constructs' Change in Behavioral Manifestation Across Development

Given the substantial changes in behavioral manifestation of psychopathology across development, it is important for measures to capture the constructs' changing behavioral manifestation, to stay developmentally relevant and construct valid. As noted by Moffitt ([1993, p. 694](#)), “Measures of antisocial behavior should be sensitive to developmental heterogeneity to tap individual differences while allowing for the emergence of new forms of antisocial behavior (e.g., automobile theft) or for the forsaking of old forms (e.g., tantrums).”

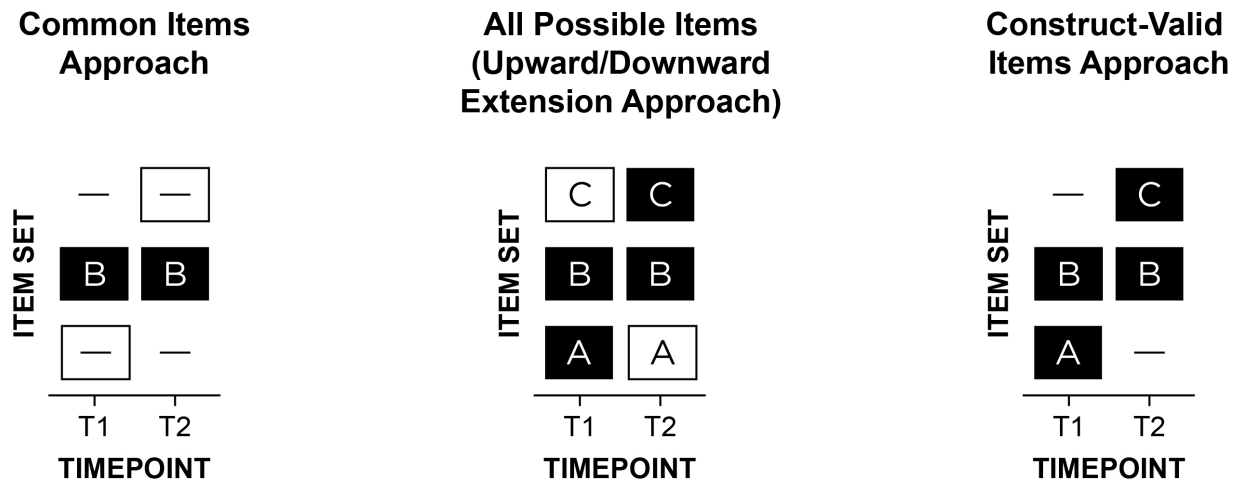
However, measures show insufficient differences in which items are assessed at which ages to account for heterotypic continuity. For example, among the most widely used measures of externalizing problems in research and practice is the Child Behavior Checklist (CBCL). However, the same version of the Child Behavior Checklist (CBCL) spans ages 6–18. Considerable development occurs during that span including key developmental transitions—from childhood to adolescence to emerging adulthood—and the CBCL does not capture such changes.

(p. 522). In addition, some disorders require that some symptoms onset before or after a particular age. For instance, in the diagnosis of ADHD, several symptoms must be present prior to age 12. In the diagnosis of CD, the symptoms “often stays out at night despite parental prohibitions” and “is often truant from school” must begin before age 13 (p. 531). In the diagnosis of ASPD, symptoms must be persistent since age 15 but have onset prior to age 15.

There are two primary ways that researchers have examined people's change in psychopathology across development (Figure 6). The “common items” approach removes items that are age specific.⁴ For example, in a study of externalizing behavior from early childhood to adolescence, a study might remove substance use items that are not be developmentally relevant at all ages. The second approach, the “upward/downward extension” approach, takes items that are valid at a given age and uses those same items across all ages, even outside the ages for which the assessment was designed and when the item may not be valid or useful—i.e., all items are assessed at all ages. For instance, the upward extension approach might take the item “disobedient to authority figures” and apply it to older individuals for whom such a behavior may no longer validly reflect externalizing behavior—and may instead reflect prosocial functions including protesting against societally unjust actions. Both the “common items” (e.g., [Odgers et al., 2008](#); [Sterba et al., 2007](#)) and “upward/downward extension” (e.g., [Briggs-Gowan et al., 2016](#); [Broeren et al., 2013](#)) approaches are widely used, likely because they result in the same items assessed across ages, but both have key problems.

As an example of the common items approach, [Odgers et al. \(2008, p. 676\)](#) examined children's development of DSM-IV symptoms of conduct disorder from ages 7 to 26 years, but they dropped age-specific symptoms (e.g., running away, staying out late) “because [these symptoms] did not cover the study's age span”. As an example of the upward extension approach, [Broeren et al. \(2013, p. 84\)](#) examined children's development of anxiety from ages 4 to 11 years; they noted, “Although the questionnaire was originally developed to measure anxiety in preschoolers, the current study also employed the scale with older children to promote uniformity in measures”. As examples of the downward extension approach, [Gornik et al. \(2023\)](#) used the CBCL 4–18 with 3-year-olds—i.e., ages younger than the assessment was designed. [Zhai et al. \(2026\)](#) used items (e.g., “steals from home, school or elsewhere”) from the Strengths and Difficulties Questionnaire for 4–10/11–17-year-olds with 3 year-olds—i.e., ages younger than the assessment was designed, despite availability of a version for 2–4-year-olds.

⁴ For simplicity and brevity, we refer to birth age as a proxy for developmental time.

Figure 6*Approaches to Longitudinal Assessment*

Note. Item set A refers to items that are construct-valid at only timepoint 1 (T1); item set B is construct-valid at both T1 and T2; item set C is construct-valid at only T2. A dash indicates that the item set was not assessed at a given timepoint. A white box indicates invalid assessment in terms of either a content gap (i.e., important missing items) or intrusion (i.e., invalid items at a given timepoint). In a study of externalizing behavior from early childhood to adulthood, “biting others” may be in item set A; “noncompliant” in B; “drug use” in C. The three approaches are: (1) common items: B at T1 and T2, (2) upward/downward extension: ABC at T1 and T2, or (3) construct-valid items: AB at T1; BC at T2. The “common items” and “upward/downward extension” approaches are by far the most widely used in the literature, even though they likely lead to inaccurate scores and thus inaccurate trajectories. (Figure reprinted from Petersen, Demko, Lee, et al. (2026), Figure 1, p. 113. Petersen, I. T., Demko, Z., Lee, W.-C., & Oleson, J. J. (2026). Studying development of psychopathology using changing measures to account for heterotypic continuity. *JAACAP Open*, 4(1), 111–123. <https://doi.org/10.1016/j.jaacop.2025.10.008>).

The common items approach yields low content validity because it does not assess all construct facets, especially age-specific manifestations. The upward/downward extension approach violates construct validity because it assesses developmentally inappropriate items.

Due to these problems, both simulation (Petersen et al., 2021) and empirical work (F. R. Chen & Jaffee, 2015; Petersen et al., 2018; Petersen, Demko, Lee, et al., 2026) have shown that the “common items” and “upward/downward extension” approaches—thus failing to account for heterotypic continuity—lead to inaccurate developmental inferences. The extent of inaccuracy may vary from small to substantial depending on the approach used, the criterion examined, and the extent to which the construct changes in manifestation (whose changes are ignored).

Nevertheless, the effects of failing to account for heterotypic continuity have been large enough to infer that (a) a sample does not show changes in level when it actually does (Petersen et al., 2018), (b) a sample shows increases when it actually decreases (Petersen et al., 2021), (c) many individuals are misclassified as persisting versus desisting in terms of their externalizing trajectory (Petersen, Demko, Lee, et al., 2026), and (d) a subgroup's externalizing behavior emerges in adulthood when it actually onsets much earlier (i.e., adolescence; F. R. Chen & Jaffee, 2015). In sum, despite the “common items” and “upward/downward extension” approaches being the most widely used for assessing individuals' development, they likely lead to inaccurate scores and thus inaccurate trajectories and ages of onset for individuals and groups.

The problem cannot be solved merely by age-norming. Generating norms from items that are invalid at a given age or from scales that have important content gaps at a given age leads to construct-invalid scores and resulting norms. Likewise, using DSM disorders or diagnoses also does not solve the problem because DSM disorders are fictive categories (Petersen, 2024) and, as described above, DSM disorder criteria do not adequately account for different symptom presentations at different ages. Moreover, diagnostic criteria naturally change with subsequent revisions of the DSM, leading people to lose interest with prior categorizations and to incomparability of diagnoses across DSM versions.

In addition, forcing use of the same measure across time naturally limits the ages a study can span, which prevents charting wider age spans. For example, if a researcher used the CBCL and only wanted to study trajectories where the same measure was assessed across ages, they would not be able to examine development from ages 5–6, when there is a change in the CBCL version (i.e., one for ages 1.5–5 and another for ages 6–18).

In addition to the behavioral expression of a construct changing across developmental time, the behavioral expression of a construct can also change across historical time. Changes in the behavioral manifestation across historical time can lead existing items and measures to become obsolete and new items to become relevant. For instance, many measures and longitudinal studies do not assess important contemporary manifestations of externalizing

behavior, including aggression on social media and cyberbullying (Nickerson & Fredrick, 2026). Changes in the behavioral manifestation across historical time provide additional reasons to change the measure by removing no-longer-valid items and by adding newer valid items.

In sum, many problems arise from using the same measure across ages when the construct changes in behavioral manifestation across development and historical time.

Do Not Allow Charting Individuals' Change Over Time When the Construct Changes in Behavioral Manifestation

Current assessments fail to implement the construct-valid item approach across development—existing measures do not assess the most construct-relevant items at particular ages and thus show insufficient differences in which items are assessed at which ages. Even in cases where the items of a measure differ across ages (e.g., CBCL: ages 1.5–5 vs. ages 6–18), the scoring systems do not link scores across ages to be on the same scale while preserving change in level. For instance, when using age-differing measures of externalizing behavior to account for changes in behavioral manifestation across development, Kjeldsen et al. (2016, p. 999) noted in their study, “Due to change in measures and rescaled variables, only relative change across classes can be interpreted and not absolute (developmental) change.” Thus, the measures do not yield scores that are comparable across time and that enable change in an individual’s level to be observed across a lengthy developmental span. In some cases, mean scores and proportion of maximum scores have been used to account for the age-differing measures having different numbers of items (Owens & Shaw, 2003; Petersen et al., 2015; Sacu et al., in press). However, such approaches assume that the age-differing measures have the same overall functioning in terms of difficulty and discrimination (described later), which is unlikely.

Age norms and standardized scores (e.g., *T*- or *z*-scores) have also been used to study development of externalizing problems (Bista et al., 2025; Korhonen et al., 2018). However, age norms and standardized scores do not allow observing individuals’ absolute change because scores have a fixed mean and variance, and they do not equal developmental understanding. Absolute change is necessary to identify true growth for an individual, and is thus important for

treatment monitoring and longitudinal growth curves. Thus, age-norming is problematic in longitudinal studies ([Moeller, 2015](#)).

It is important for scores across ages to be on the same apples-to-apples scale so the scores can be compared across ages, which would be valuable even in a cross-sectional study that spans a wide age range, as might be found in an epidemiological study. There is a misperception that no practical approach exists that can handle different measures across ages or time. Consider the following statement by LeBlanc ([2021, p. 15](#)), “The theoretical position is that a form of antisocial behavior, for example, fraud, has to be measured by distinct behaviors with adolescents and adults, in summary adapted to their age. Nevertheless, different scale composition is an integral part of the common concept of fraud. To our knowledge there is no recognized analytical procedure to assess such a construct and the life cycle validity of heteromorphous measures.”

However, considerable research in educational psychology links scores across age-differing measures to be on the same scale. For instance, standardized tests are frequently administered in schools to determine the extent of growth across grades. Because of substantial changes in ability level across grades, the same math and reading items cannot be administered at all grades. Otherwise, respondents would receive items that are too easy (thus evoking boredom) or too difficult (thus evoking frustration), relative to their ability. To address this, the test developers write the tests so that some items are in common across adjacent grades—used for linking the scores across grades onto the same scale—along with items that are unique to a particular grade. Then, the researchers apply widely used methods to link the scores from the different measures onto the same scale so they can observe individuals’ growth across grades. The same general methods are also used to place people’s scores on the same scale for standardized tests like the ACT, SAT, GRE, MCAT, LSAT, etc., despite respondents receiving different item content.

For examples of studies applying various linking methods to academic and cognitive skills, see McArdle et al. ([2009](#)), McArdle and Grimm ([2011](#)), Murayama et al. ([2013](#)), Oleson et al. ([2016](#)), Petscher et al. ([2018](#)), Ward et al. ([2022](#)), and Weeks ([2018](#)). However, the use of linking

methods is rare in studies of social and behavioral development broadly and psychopathology specifically. For studies that have applied such methods in studying social and behavioral development, see studies on self-concept (Hancock & Buehl, 2008), neuroticism (Lai, 2023), self-regulation (Hosch et al., 2022), underactivity/overactivity (and their subdimensions; McDermott et al., 2013), thought-disordered problems (Petersen, Demko, Lee, et al., 2026), internalizing problems (Harris et al., 2025; Petersen et al., 2018; Petersen, Demko, Lee, et al., 2026; Tyrell et al., 2019), and externalizing problems (S. M. Chen & Bauer, in press a, in press b; Harris et al., 2025; Petersen, Demko, Lee, et al., 2026; Petersen & LeBeau, 2021, 2022).

A failure to apply linking methods has consistently been shown to yield less accurate trajectories—at the person and group level (S. M. Chen & Bauer, in press a; Petersen et al., 2018; Petersen et al., 2021; Petersen, Demko, Lee, et al., 2026). Thus, it is imperative that studies link scores across age-differing measures of psychopathology. However, psychopathology assessments do not include scoring tools for linking scores across age- (or rater-) differing versions. The lack of scoring tools that link scores from age-differing psychopathology assessments leads researchers to restrict the age range of study. For instance, in their study, Bista et al. (2025, p. 179) noted “A different version of the questionnaire (CBCL/2–3), designed for 2–3-year-olds, was used for the year 2 follow-up [compared to the CBCL/4–18 assessed from ages 5–17]. Therefore, we excluded age 2 from our trajectory analysis.”

Not Well-Suited for Capturing the Full Range of Individual Differences

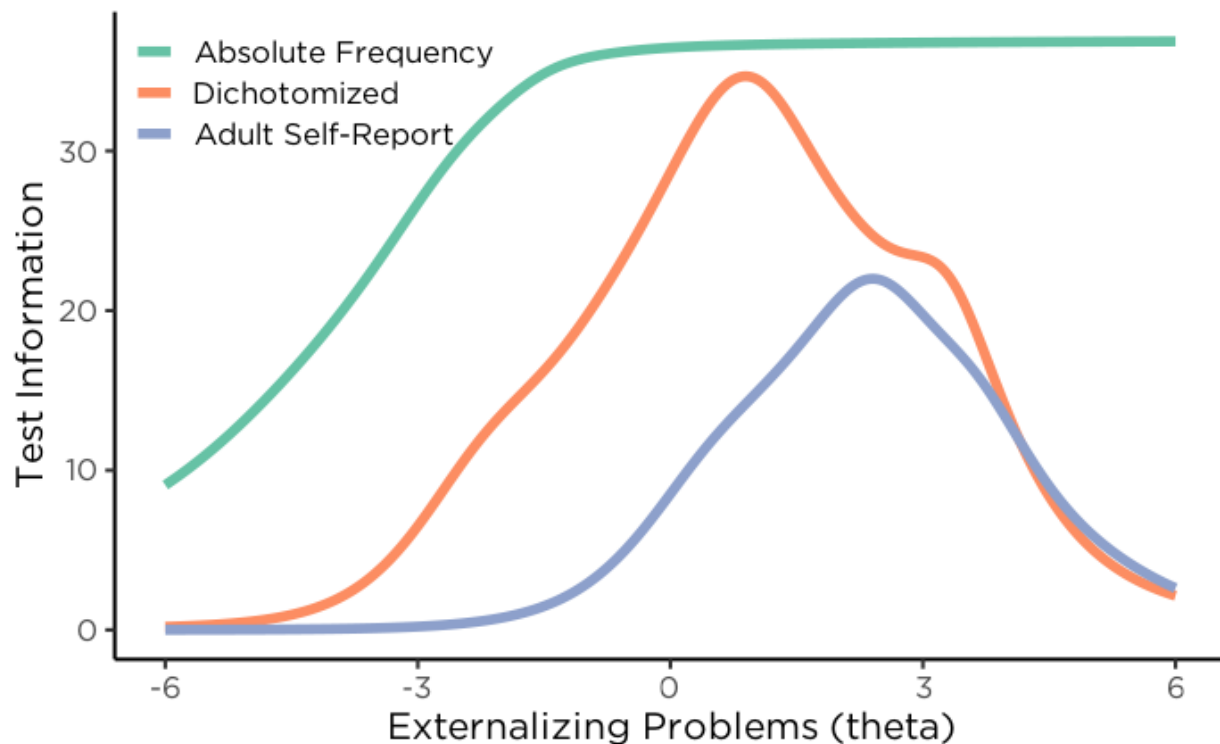
Many measures use coarse, subjective response scales such as Likert-type response scales (e.g., “rarely,” “sometimes,” “often,” “very often”) leading to cultural bias (Paalman et al., 2013; Schaeffer, 1991; Schwarz, 1999) and insensitivity to change. Items with few discrete response categories tend to show less sensitivity to the full range of individual differences (Elliott & Ageton, 1980) and less sensitivity to change (i.e., responsiveness; Grant et al., 1999) compared to items with more response options that have greater variability in scores. In general, items with more response options yield more information (i.e., greater reduction of uncertainty in estimates of a person’s construct level; though there are diminishing returns when using more than six

response options; [Culpepper, 2013](#); [Simms et al., 2019](#)). Indeed, as depicted in Figure 7, items assessing absolute frequency have been shown to yield more information than dichotomized versions of the same items and than the Achenbach Adult Self-Report, which uses Likert-type response options. In addition, if a respondent implicitly uses the target's same-aged peers as the normative reference group for determining the relative frequency of a behavior, this could have the implicit effect of resulting in age-normed scores, which works against the goal of detecting change. Moreover, Likert-type scales tend to focus on frequency of misbehavior and ignore intensity and severity. It can be helpful to explicitly assess—and distinguish—frequency versus problematic impact of behavior. Instead of assessing relative frequency, researchers advise assessing absolute frequency of behavior ([Burns et al., 2001](#); [Schwarz, 1999](#))—in addition to other aspects of behavior including onset, timing, duration, intensity, function, and resulting impairment.

In addition, the measures focus on extreme or clinical-range behaviors; scores are thus positively skewed with floor effects and have unacceptably low reliability (including the CBCL; [Kaat et al., 2019](#); [Pavlovich et al., 2026](#); [Tiego et al., 2023](#)), making the CBCL and existing measures unsuitable for dimensional research ([Pavlovich et al., 2026](#); [Tiego et al., 2023](#)). Considerable evidence indicates that externalizing problems (and psychopathology, more generally; [Haslam et al., 2012](#); [Haslam et al., 2020](#); [Kotov et al., 2017](#); [Krueger & DeYoung, 2016](#); [Krueger & Markon, 2011](#); [Krueger & Piasecki, 2002](#); [Markon et al., 2011](#); [Wright et al., 2013](#)) and their trajectories ([Burt, 2012](#); [Fairchild et al., 2013](#); [Walters, 2011, 2012, 2015](#); [Walters & Ruscio, 2013](#)) and risk factors ([Walters, 2014](#)) are best conceptualized dimensionally (or multi-dimensionally; [Bolhuis et al., 2017](#); [Wakschlag et al., 2014](#)), not categorically ([Barry et al., 2013](#); [Coghill & Sonuga-Barke, 2012](#); [Forbes et al., 2016](#); [Kliem et al., 2018](#); [Kliem et al., 2022](#); [Krueger et al., 2004](#); [Krueger et al., 2005](#); [Krueger et al., 2007](#); [Krueger et al., 2021](#); [Markon & Krueger, 2005](#); [Sellbom, 2016](#); [Walton et al., 2011](#)). Thus, the existing measures are impoverished for studying the full range of individual differences on the externalizing spectrum, which prevents subthreshold children who are at risk for impairment from being identified. Better identification of subthreshold externalizing is critical, in both research and practice, for identifying at-risk

Figure 7

Test Information Functions of Absolute Frequency Versus Dichotomized Versus Adult Self-Report Items: Test Information as a Function of Theta (θ)



Note. “Absolute Frequency” refers to count items; “Dichotomized” refers to dichotomized versions of the count items; “Adult Self-Report” refers to items from the Achenbach Adult Self-Report. Theta represents the person’s level on the latent externalizing problems factor. The figure demonstrates that items assessing absolute frequency yield more information (i.e., measurement precision) than dichotomized versions of the same items and the Likert-type response options of the Adult Self-Report. (Figure reprinted from Petersen, Demko, Doebler, et al. (2026), Figure 4, p. 303. Petersen, I. T., Demko, Z., Doebler, P., Sabel, L., Oleson, J. J., & Krueger, R. F. (2026). How often is “often”? Improving assessment of the externalizing spectrum using absolute frequency. *Psychological Assessment*, 38(4), 295–306. <https://doi.org/10.1037/pas0001441>).

individuals in the early-stage trajectory of psychopathology. As noted by Pavlovich et al. (2026, p. 1), “These findings...highlight the need for measures explicitly developed and validated for dimensional psychopathology in population-scale research.”.

To provide better phenotypic resolution, and to better identify biological processes underlying psychopathology, Tiego et al. (2023) argue for including content covering the full range, including items assessing the adaptive end of the continuum. Doing so would help move the field away from a focus of symptoms (and absence of symptoms) and toward consideration of the full range of functioning, including thriving. As they note, low scores on clinical measures are typically intended to assess absence of pathological functioning rather than the opposite end of the latent continuum—i.e., the presence of adaptive functioning. For instance, it can be valuable to consider “positive opposites”—i.e., positive behaviors that compete with and are (simultaneously) incompatible with the particular antisocial behavior. For instance, in a study of externalizing behavior, it would be valuable to consider prosocial behavior (e.g., kindness, helping others, sharing, dependability, compliance, accountability). Indeed, in a machine learning study of children’s conduct problems in the ABCD study, prosocial behavior was the strongest predictor (Berluti et al., 2026). Moreover, measures of externalizing behavior tend to focus on just global scores or just one or a few subdimensions. For better phenotypic resolution, instruments should assess a wide range of phenotypes to better capture the full range of externalizing behaviors, which will be important to distinguish various forms of behavior—e.g., physical versus verbal versus destructive versus relational aggression—and their functions—e.g., proactive versus reactive.

Based on our content analysis of over 270 measures of externalizing behavior, we identified a wide range of facets that would be important for measures of externalizing behavior to capture (see Table 1). We also identified potential positive opposites of each facet.

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Apathy/Indolence/Disengagement	Unconcerned about achievement, low motivation, low concern about how one's abilities or actions are evaluated, specific to contexts in which others have expectations for performance (e.g., school and work), specifically low effort rather than performance	Diligence/Achievement motivation
Attention seeking/Exhibitionism	Exaggerated or dramatic expressions, interrupting others, inappropriate or provocative behaviors, provoking reactions	Humility/Appropriate self-expression

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Behavioral addiction	Impulsive/compulsive engagement in rewarding non-substance behaviors (e.g., gambling, video games, risky sex, pornography, food, the internet, mobile devices, shopping)	Self-regulated engagement
Blame externalization	Attribute cause of problems or failures to external factors, falsely claiming victimhood, blaming others	Personal responsibility/Accountability
Boredom proneness	Expressing boredom often, difficulty maintaining interest, lack of engagement	Curiosity/Sustained engagement

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Callous aggression	Inflicting harm without feeling remorse or concern, usually proactive/intentional	Compassionate protection/Benevolence
Callousness/Disregard for others/Selfishness/Egocentricity	Emotional detachment, indifference to well-being of others	Empathy/Compassion/Concern for others/Helping others/Sharing
Deceitfulness/Fraud/Lying/Cheating/ Keeping secrets/Stealthiness/Exploitation	Deceitfulness/ hidden, especially to get an advantage (e.g., lying, withholding information, manipulation, making false promises)	Honesty/Integrity/Fairness
Deviant peer affiliation	Affiliating with peers who engage in deviant or delinquent behaviors	Prosocial peer affiliations

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Destructive aggression/Destruction of property	Destruction of property, self, or others, objects	Constructiveness/Care for property
Distractibility/Inattention/Attention dysregulation/Forgetfulness/Daydreaming	Difficulty maintaining focus, attention, or concentration, frequent attention shifts	Attentional control/Sustained attention
Domineering/Bossiness	Asserting control, power, and authority in an overbearing or oppressive manner	Collaborativeness
Excitement seeking/Risk taking/Sensation seeking	Tendency to impulsively pursue new and different sensations, especially ones that are intense, exhilarating, thrilling, unsafe, etc. (e.g., risk taking, preference for novelty, intensity)	Prudent exploration/Measured novelty seeking

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Grandiosity/Overconfidence	Exaggerated sense of self-importance, sense of superiority or specialness	Realistic self-confidence
Hostility/Rudeness	Hostile behavior, unfriendliness, resentment, antagonism, ranging from mild irritation to intense rage, interpret others' behavior as hostile	Warmth/Friendliness/Kindness
Hyperactivity	High levels of physical activity, restlessness, difficulty engaging in quiet activities	Behavioral regulation
Impulsivity/Disinhibition/Impatient urgency	Tendency to act without thinking, disinhibited engagement	Planful control

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Irresponsibility/Unreliability/Disorganization	Failure to meet obligations, avoiding addressing difficulties, inconsistency, lack of organization and planning	Responsibility/Conscientiousness/Dependability/Reliability
Manipulativeness/Psychological aggression	Tendency to influence or control others to achieve one's own goals or desires, e.g., playing the victim, flattery, gaslighting, guilt-tripping, psychological aggression	Sincerity/Authenticity
Mistrust/Suspiciousness/Paranoia	Lack of confidence in the reliability, intentions of others, difficulty forming close relationships with others, difficulty forgiving	Interpersonal trust (appropriately calibrated)

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Oppositionality/Stubbornness	Persistent and defiant attitude, tendency to challenge authority, refusal to comply, argumentativeness (could include while complying), etc.	Cooperativeness
Physical aggression	Use of physical force or violence to harm others or their property	Respect for others' bodies/Helping others
Punishment insensitivity	Does not change behavior after punishment, high tolerance for negative outcomes	Responsive to consequences
Rebelliousness/Rule-breaking/Noncompliance	Doing things differently than one is asked to do (e.g., breaking house rules)	Compliance/Civic responsibility

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Relational aggression	Aggression intended to harm others through deliberate manipulation of their social standing and relationships (e.g., exclusion, speaking negatively of another, etc.)	Inclusion
Sexual aggression	Engaging in sexual activity with someone who does not or cannot consent to that behavior, or threatening to do so (e.g., harassment)	Respect for consent
Impulsive sexual behavior/Risky sexual behavior	Engaging in risky sexual activity (e.g., sex without contraceptives)	Responsible sexual decision-making
(Harmful) Substance Use/Problems	Harmful use of any drugs (e.g., alcohol, tobacco, illegal drugs)	Healthy choices around substance use

Table 1*Externalizing Facets Identified in Content Analysis of Externalizing Measures.*

Facet	Definition	Potential Positive Opposite
Theft/Stealing	Unauthorized taking of someone else's property without their knowledge or consent	Respect for people's possessions
Verbal Aggression	Communication with an intention to harm an individual through words, tone, or manner	Respectful communication

Note. Transdiagnostic facets of psychopathology, which encompass externalizing problems, include irritability/reactive anger, unemotionality, sleep problems, social problems, learning problems, and low functional independence.

Source: [Analyses](#)

Because existing measures do not capture the full range of individual differences, this naturally further limits the measures' sensitivity to detect change (i.e., responsiveness), which is crucial for clinical utility and for longitudinal research. Moreover, it is important for measures not to show restriction of range owing to ceiling or floor effects.

Not Well-Suited for Capturing Individuals' Change Over Time

In sum, although they have been helpful, existing measures are poorly suited to classify individuals' development of psychopathology across a lengthy span, which is necessary for tracking individuals' change and identifying mechanisms in the development of psychopathology. As we described above, existing assessments of externalizing behavior—and of other forms of

psychopathology—have key limitations. To address this concern, the field will need to develop better assessments. In addition, there are problems with how assessments are used to examine development—studies have primarily used problematic approaches including the common items approach or the upward/downward extension approach across ages, leading to inaccurate scores and resulting trajectories and developmental inferences. Moreover, studies have largely failed to place scores on a comparable scale across development. Researchers studying psychopathology development are doing their best to grapple with developmental equivalence in the ways they know how, such as by dropping items, using items outside the age range of validation, restricting the age range of analysis (or data collection) to the ages when the same measure was (or would be) used, using diagnostic categories, averaging scores, and age-norming. However, it is important for the field to advance its awareness of stronger approaches for placing scores on a comparable scale across development. To address these concerns, we propose a potential solution to account for heterotypic continuity and to allow charting individuals' change across the lifespan: a developmental scaling framework.

Solution: A Developmental Scaling Framework

A developmental scaling framework informs both assessment—including item selection—and linking of scores.

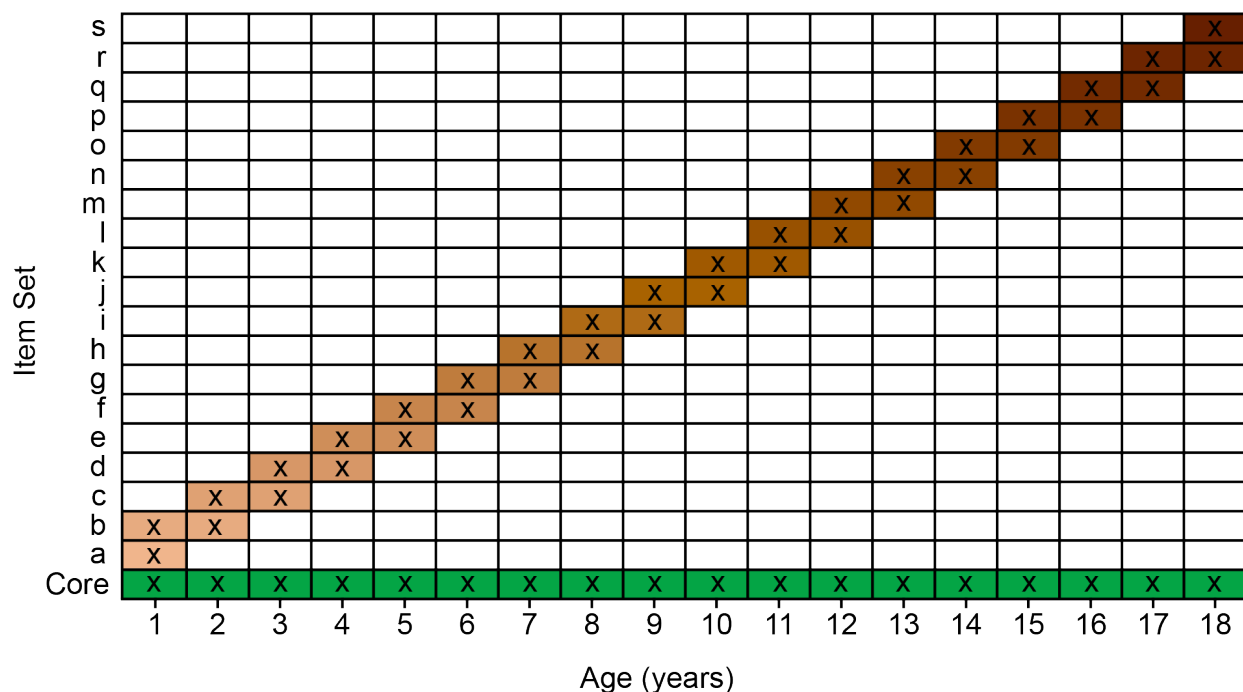
Item Selection

When a construct changes in behavioral manifestation across development, which we would argue is the norm rather than the exception when considering constructs across the lifespan, it is essential to account for such changes in item selection. Thus, it is important to use the construct-valid items at a given age, even if that leads to using different measures across ages. A conceptual depiction of an example construct-valid item design spanning infancy to adulthood is in Figure 8. In the construct-valid item design, some items represent the core of the phenotype and are assessed across all ages, whereas other items are age-specific. Despite some age-specific items, some age-common items are assessed at adjacent ages, which help serve as anchor items for linking the scores across the adjacent ages. The use of overlapping yet distinct items is widely

used in educational testing to help link individuals' scores across grades, and is called vertical scaling because the item sets appear to form a line moving upward (see Figure 8). For example, for a standardized test of math ability, kindergartners may receive items assessing counting, 1st graders may receive items assessing counting and addition, 2nd graders may receive items assessing addition and subtraction, and so on.

Figure 8

Conceptual Depiction of a Construct-Valid Item Design from Infancy to Adulthood



Note. The age(s) at which an item set is assessed are represented by the letter x and a colored box. Some items (“Core”, in green) represent the core of the phenotype and are assessed across all ages. Some items (in varying shades of brown) are age-specific and are assessed at only one or a few ages. However, some of the items assessed at a given age are shared with those assessed at adjacent ages, to facilitate linking the scores across ages. For instance, item set f can be used to help link scores from age 6 to those at age 5, and item set g can be used to help link scores from age 6 to those at age 7. Hypothetically, items could be assessed for as many or as few years as desired—as long as the items are construct valid for those ages and there enough age-common items between adjacent ages for linking scores across ages.

Linking of Scores

After the age-differing measures have been assessed, the scores from the age-differing measures can be linked to be on the same scale. To perform developmental scaling, a variety of

approaches can be used, including partial measurement invariance, the alignment method, moderated nonlinear factor analysis (MNLFA), and item response theory (IRT). These are all approaches to harmonization—i.e., linking scores from different measures onto the same scale so they can be meaningfully compared. Harmonization is also used for integrative data analysis, to pool data from different measures in various studies ([Curran et al., 2008](#); [McDaniel et al., 2023](#)). The aforementioned approaches to harmonization attempt to estimate latent variables—using either confirmatory factor analysis (CFA) or IRT—on a common scale across ages. They do so by trying to align the item parameters for at least a subset of items. To be able to chart individuals' change over time, it is helpful to ensure that the scores are on the same number line across time—that is, they have the same scale and location across ages. Scale refers to the size of a unit on the number line. If scores have the same scale across ages, a one-unit difference represents the same amount of the construct at each age. Location refers to the origin or zero point on the number line. If scores have the same location across ages, a score of zero corresponds to the same position on the construct across ages. Collectively, if scores have the same scale and location across ages, the measure is constantly calibrated across ages, and the same score corresponds to the same level on the construct across ages. CFA and IRT each have item parameters corresponding to scale and location.

In CFA, scale is informed by an item's factor loading, and location is informed by an item's intercept. A factor loading reflects the strength of association between the item and the latent factor, with larger loadings indicating stronger associations. For instance, an item assessing the extent to which a child hits others would have a stronger factor loading on an externalizing factor than an item assessing the extent to which a child likes ice cream. The intercept is the expected score on the item when the latent factor equals zero. When the latent factor is centered to have a mean of zero, the intercept represents the expected score on the item for someone at the mean level of the construct. Items with lower intercepts tend to be endorsed by individuals with higher levels of the latent factor and therefore often reflect more severe behaviors. For example, an item assessing the extent to which a child sets fires would be expected to have a lower intercept

than an item assessing the extent to which a child argues.

In IRT, scale is informed by an item's discrimination, and location is informed by an item's difficulty. An item's discrimination reflects how strongly the item differentiates individuals at various levels of the construct, with higher discrimination indicating a stronger association with the latent factor. In a two-parameter IRT model, an item's difficulty reflects the level on the latent factor where there is a 50% probability of endorsing the item. Items with higher difficulty require a higher level of the latent factor to have a 50% probability of endorsing the item and therefore represent more severe behaviors.

Partial Invariance

The partial invariance method can be used to link scores across different measures (Tyrell et al., 2019) and can be paired with second-order growth functions to examine individuals' change over time (Hancock & Buehl, 2008). The partial invariance approach attempts to establish partial strong longitudinal factorial invariance⁵—i.e., invariant intercepts and factor loadings across ages for a subset of items to set the latent factor across ages on a comparable scale, with some items' intercepts and/or factor loadings allowed to differ across ages. The researcher first identifies which items show invariant factor loadings and/or intercepts and which items show noninvariant factor loadings or intercepts. The researcher then constrains the factor loadings across time for any items with invariant factor loadings, to serve as anchor items and establish an equivalent scale of the measures across time, and allows the factor loadings to differ across time for the items with noninvariant factor loadings. In addition, the researcher constrains the intercepts across time for any items with invariant intercepts, to serve as anchor items and establish an equivalent location of the measures across time, and allows the intercepts to differ across time for the items with noninvariant intercepts. The anchor items thus equate the latent factor over time, preventing drift in the scale calibration and rendering scores comparable across time (S. M. Chen & Bauer, in press b). Then, the researcher estimates second-order growth functions (e.g., intercept and slope) of the first-order latent factor across time. The method can allow different indicators across ages,

⁵ Strong invariance is also called “scalar invariance”.

either by dropping or adding indicators or by using item parcels composed of differing items across ages (Tyrell et al., 2019).

However, the partial invariance approach requires identifying—and accurately specifying—invariant items and any noninvariant items, which can be both inefficient and error-prone (S. M. Chen & Bauer, in press b; Lai, 2023). Moreover, the partial invariance approach tends not to work as well when a large proportion of items are noninvariant (Lai, 2023), there are many timepoints, there is considerable variability in the numbers of responses provided to items such as might occur due to missing data or when not all items are assessed at all timepoints, or the researcher wants to consider factorial invariance or differential item functioning (DIF) as a function of multiple variables or continuous variables (S. M. Chen & Bauer, in press b).

Alignment Optimization

The alignment optimization method (Muthén & Asparouhov, 2014) has been used to link scores across time (Lai, 2023). The alignment optimization method within confirmatory factor analysis attempts to identify a model with approximate longitudinal factorial invariance—that is, to retain any instances of large noninvariance across ages and to keep other parameters approximately invariant across ages—to adjust for violations of invariance. The approach uses a component loss function to align the item parameters across ages so that the scores on the latent factors are on the same scale. Unlike the partial invariance method, the alignment method does not require a priori identification of any noninvariant items (Lai, 2023).

Moderated Nonlinear Factor Analysis

MNLFA (Bauer, 2017) has also been used to link scores across different measures (Curran et al., 2014), including across ages (S. M. Chen & Bauer, in press b, in press a; Student, 2025). Unlike the partial invariance approach, MNLFA represents the measurement model in continuous time (rather than discrete time). MNLFA allows simultaneous estimation of growth functions while estimating and accounting for DIF effects. It allows estimating DIF as a function of time and as a function of multiple variables, including categorical and continuous variables. Bayesian approaches can be used to apply MNLFA. A key benefit of Bayesian approaches is the use of

regularizing priors that allow estimating the model without having to identify noninvariant/DIF items and anchor items in advance (S. M. Chen & Bauer, *in press b*, *in press a*). Using regularization, unimportant DIF effects are shrunk toward zero—while retaining larger DIF effects—thus identifying potential anchor items without having to specify them in advance (S. M. Chen & Bauer, *in press b*). In addition, Bayesian approaches are often better suited to handling small sample sizes and data with sparse responses (S. M. Chen & Bauer, *in press b*). Using a simulation, Chen and Bauer (*in press a*) found that MNLFA estimated change more accurately than traditional first- or second-order growth curve models.

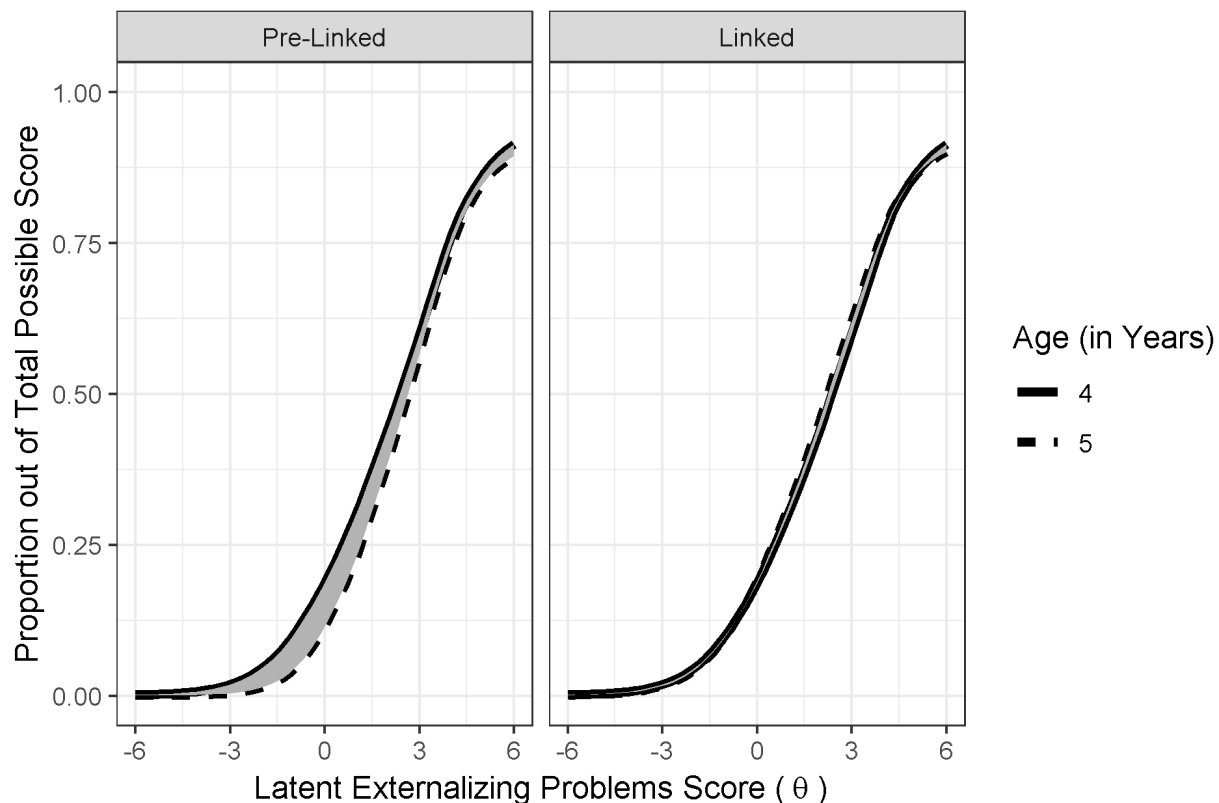
Item Response Theory

IRT can also be used to link scores across different measures (Kolen & Brennan, 2014). The IRT approach to developmental scaling is depicted in Figure 9. The IRT approach to developmental scaling identifies scaling parameters that minimize the differences in the probability of a person endorsing the age-common items across two measures, at a given construct level. That is, IRT links measures' scales based on the difficulty and discrimination of the age-common items. In developmental scaling, scores on the construct-valid items at the reference age set the scale; the age-common items adjusts subsequent scores to that scale, and all construct-valid items (i.e., both age-common and age-unique items) at a given timepoint are used to estimate each person's score on that scale. Thus, the age-common items are used to determine the general form of change on an identical scale, but all developmentally relevant, construct-valid items are used to estimate each person's construct level on this scale. Traditionally IRT represents the measurement model in discrete time. Like MNFLA, however, IRT (particularly in the mixed model formulation) can represent the measurement model in continuous time and allows testing DIF over time and as a function of multiple variables and continuous variables. Moreover, like MNLFA, it allows simultaneous estimation of growth functions while estimating and accounting for DIF effects. In addition, like MNFLA, IRT allows Bayesian estimation and regularizing priors to identify anchor items without having to specify them in advance.

Nevertheless, for many of these approaches, at least one anchor item is necessary to link

Figure 9

Example of Linking Scores From Measures Across Ages Using the Item Response Theory Approach to Developmental Scaling



Note. The figure illustrates the effect of linking the latent externalizing problems scores, θ , across ages, using developmental scaling. The left panel illustrates the test characteristic curves representing the model-implied proportion out of total possible scores across the latent externalizing problems score at age 4 and 5, before the linking process. The right panel illustrates the test characteristic curves after the linking process. The shading between the age 4 and age 5 test characteristic curves represents differences between the two test characteristic curves in terms of discrimination and/or severity, where larger differences reflect scores that are less comparable. Linking minimizes differences between the discrimination and severity of the age-common items. Discrimination is depicted by the steepness of the slope at the inflection point of the test characteristic curve. Severity is represented by the value on the x-axis at the inflection point of the test characteristic curve. The left panel indicates that the externalizing problem items showed higher severity (i.e., its inflection point is further right) and somewhat stronger discrimination (i.e., its slope is steeper) at age 5 than at age 4. The right panel shows considerably smaller differences between the two test characteristic curves, which provides empirical evidence that the linking successfully placed the latent externalizing problem scores across age on a more comparable scale (i.e., more similar discrimination and severity of the age-common items). (Figure reprinted from Petersen and LeBeau (2021), Figure 1, p. 74. Petersen, I. T., & LeBeau, B. (2021). Language ability in the development of externalizing behavior problems in childhood. *Journal of Educational Psychology*, 113(1), 68–85. <https://doi.org/10.1037/edu0000461>).

the scale of the latent factor across timepoints ([S. M. Chen & Bauer, in press b](#)). Moreover, when treating time continuously, the functional form of any DIF as a function of time must be correctly specified ([S. M. Chen & Bauer, in press b](#)).

Other Key Challenges to Address

In addition to linking scores across time to be on the same scale, to study the development of psychopathology across the lifespan, there are additional key challenges to address, including considering and/or addressing potential confounds of change, factorial invariance, integration across multiple sources, and obsolescence of items.

Potential Confounds of Change

It is important for longitudinal research to consider and, to the extent possible, account for potential confounds of change, such as practice effects, period and cohort effects, measurement error, selective attrition, and changes in the meaning of the scores ([Hertzog & Nesselroade, 2003](#)). By confound of change, we mean that such effects can make it appear that people have changed in their level on the construct when, in reality, they have not changed (or vice versa).

Practice/Retest Effects

Practice effects (also called retest effects) occur to the extent that a person's score changes due to repeated assessment rather than due to changes in their construct level. For instance, if researchers intend to study people's intelligence across development and they use the same measure and stimuli at each measurement occasion, it is possible that people may perform better over time because they are learning the task and thus improving their score—even if their intelligence has not changed. In a longitudinal study of externalizing problems using repeated surveys, retest effects might be most likely to be detected in the form of decreases for a person across ages, which may reflect fatigue in reporting ([Averdijk, 2014](#)). Lower ratings for an individual across time may be especially likely if the surveys include branching logic such that the respondent learns that additional items appear whenever they endorse an item (and that those additional items would not have appeared had they not endorsed the item). Ways to reduce retest effects include changing the measure and/or stimuli ([Tyrell et al., 2019](#)) and increasing the retest

interval. Thus, an added benefit of accounting for heterotypic continuity (by changing the measure across development to align with changes in the construct manifestation) is the reduction of retest effects. A way to account for retest effects is through the use of accelerated longitudinal (i.e., cross-sequential) designs that can separately estimate the effects of age versus timepoint because, unlike in a single-cohort longitudinal design, age and timepoint are decoupled ([Ferrer et al., 2004](#)).

Period and Cohort Effects

Researchers are commonly interested in how people develop as a function of age. Age refers to a person's chronological age at the time of measurement. Age is a proxy for development—a stand-in for the myriad ways in which one's biology and social roles—and one's corresponding employment, income, and family structure—change as people age ([Rohrer, 2025](#)). However, apparent age-related differences in a cross-sectional or longitudinal study could reflect period or cohort effects instead of development. Thus, it is important to consider and, to the extent possible, account for period and cohort effects.

Period refers to the historical time at which measurement occurs. If people are assessed during a given time period (e.g., during a war, recession, pandemic, or stressor), their scores could be influenced by the historical time period. The extent to which scores are attributable to the time period (e.g., the economic situation or prevailing standard of living) in which they are assessed is called a period effect.

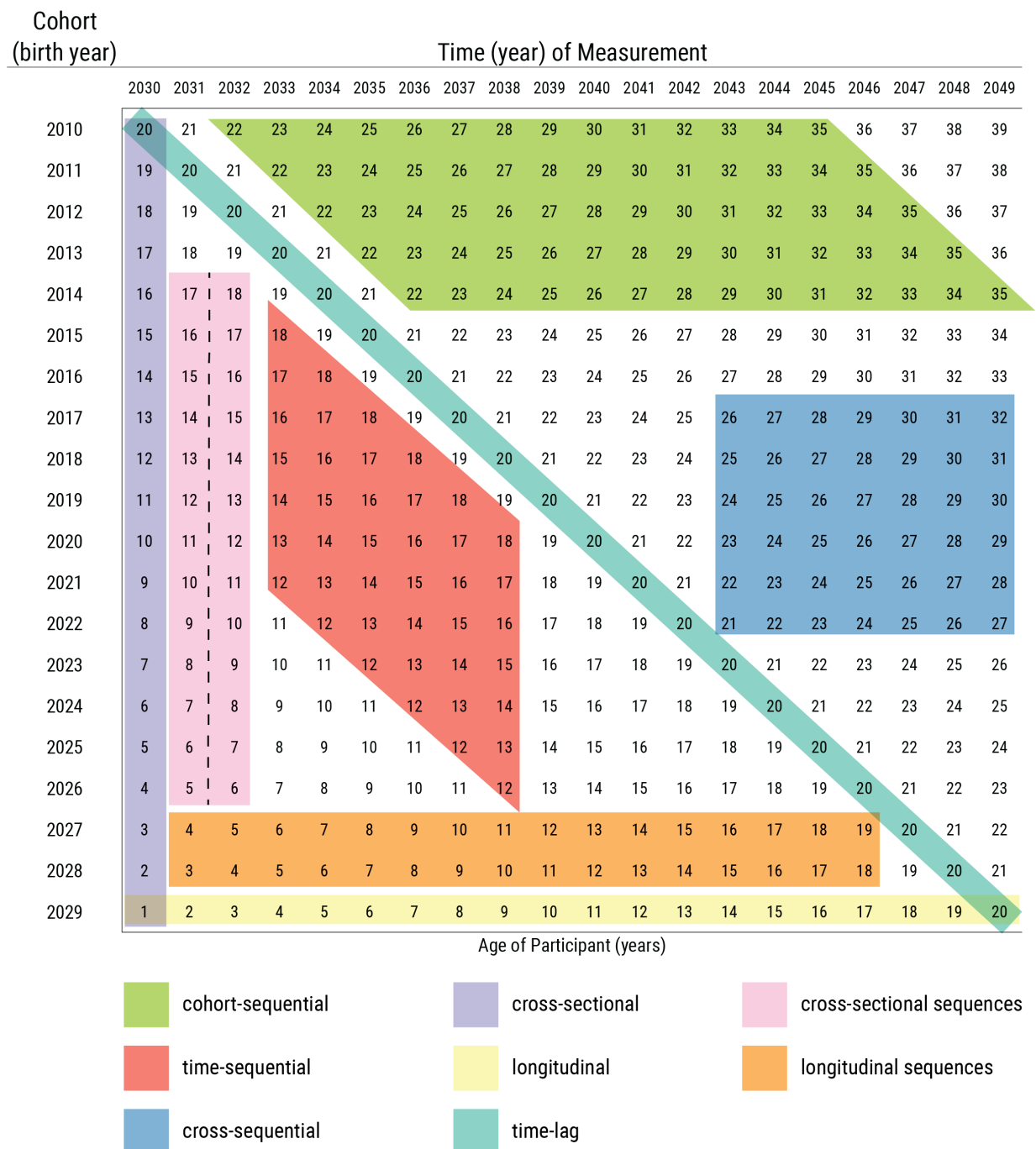
Cohort refers to a group of people who experience similar events or historical contexts at similar ages. Cohort commonly refers to birth cohort, which represents the group of people who were born around the same time. If a group of people were born around the same time (e.g., in 1940), their scores may differ from those of people born in other eras because they experienced different historical, social, educational, economic, or technological contexts during development. The extent to which scores are attributable to an individual's birth cohort is called a cohort effect—that is, “the (enduring, constant) effects of having been a certain age at a certain point in historic time.” ([Rohrer, 2025, p. 2](#)). Period and cohort—like age—are proxy variables for other, causal processes.

Part of the challenge of estimating the unique contributions of age, period, and cohort is that they are deterministically related. If you know two of these (e.g., age and period), you can determine the third (e.g., cohort). That is, $\text{age} = \text{period} - \text{cohort}$.

Various types of cross-sectional and longitudinal designs are depicted in Figure 10. In a cross-sectional study, any age-related differences are confounded with cohort differences. Thus, age-related differences in a cross-sectional study may not reflect development. In a single-cohort longitudinal design, any observed changes with age could be due to period effects (rather than developmental change associated with aging) or could be specific to that cohort. To combat these limitations, studies may employ longitudinal sequential designs that incorporate multiple cohorts, such as time-sequential, cross-sequential, or cohort-sequential designs. For instance, a cross-sequential study (also called accelerated longitudinal study) starts as a cross-sectional study with participants from multiple cohorts, and then follows the participants longitudinally. In a cohort-sequential design, multiple cohorts enter the study at the same age (e.g., age 22) and are then followed longitudinally (typically across the same duration and ages), like starting a longitudinal study at the same age over and over again. A time-sequential design is an extension of a time-lag design. A time-lag design assesses multiple cohorts at the same age (e.g., age 12). A time-sequential design adds age groups to a time-lag design, repeatedly assessing the same age range (e.g., ages 12–18), with some participants followed longitudinally.

Longitudinal sequential designs provide greater leverage for distinguishing age, period, and cohort influences by collecting information from multiple cohorts at multiple periods and ages. Nevertheless, they do not eliminate the confounding between age, period, and cohort (Rohrer, 2025). No research design can fully disentangle age, period, and cohort effects. Confidence that observed age-related differences reflect developmental change increases when there is converging evidence across multiple research designs. However, in a given study, there may be ways to account for period or cohort effects. For instance, in an accelerated longitudinal study, it would be valuable to evaluate the extent to which the differing cohorts are comparable—i.e., that they show similar scores at the same ages. Moreover, there is an entire set

Figure 10
Research Designs by Time of Measurement and Cohort



Note. Values in the cells are ages of the participants. Dashed line indicates different participants were assessed at each time of measurement. (Figure reprinted from Petersen (2026), Petersen, I. T. (2026). *Principles of psychological assessment: With applied examples in R*. University of Iowa Libraries. <https://doi.org/10.25820/work.007199>). Although the depictions of cohort-sequential and cross-sequential designs differ from the depictions by Schaie (2005) (i.e., they are reversed), they are consistent with contemporary definitions of these designs (Little, 2024; Masche & Dulmen, 2004; Whitbourne, 2019).

of analytic approaches, known as age-period-cohort analysis ([Yang & Land, 2013](#)), whose goal is to disentangle the effects of age, period, and cohort. Rohrer ([2025](#)) provides additional suggestions for considering age, period, and cohort effects.

Measurement Error

Measurement error is another potential confound of change that can make it look like change has occurred when, in reality, change has not occurred. Consider the difference between scores at two measurement occasions, a so called “difference score,” which is a common way of examining change. Difference scores are notoriously unreliable. Difference scores tend to be lower in reliability than each of the indices that comprise it, especially when the two indexes are correlated because differences depend on the reliability of both indices in the subtraction ([Revelle & Condon, 2019](#)). To be reliable, difference scores require high reliability of the individual indices compared to the correlation between them. That is, the more two things are the same thing, the more likely that subtracting one from the other leaves measurement error rather than construct variance. Ways of accounting for (random) measurement error include (a) combining multiple measures in estimation of a latent variable and (b) using growth curve models with at least three timepoints to estimate model-implied trajectories and reduce the influence of occasion-specific random error.

In general, the inference of change is strengthened to the extent that: (a) the magnitude of the difference between the scores at the timepoints is large (i.e., a large effect size), (b) the measurement error (unreliability) at each timepoint is small, (c) the measure has the same meaning and is on a comparable scale at each timepoint, and (d) evidence suggests that the differences across time are not likely to be due to potential confounds of change such as practice effects, cohort effects, period effects, or selective attrition.

Selective Missingness and Attrition

Another potential confound of change is selective (or systematic) missingness or attrition. If people who drop out of a longitudinal study differ in systematic ways from those who are retained, inferences about growth patterns and about associations with growth parameters can be

biased. Consider an intervention study with two groups: a treatment group that received medication and a control group that received a placebo. The control group showed a 30% reduction in externalizing problems, on average. In the treatment group, some individuals showed a reduction in externalizing problems whereas others dropped out of the study due to side effects or worsening behavior. If we analyzed only those who were followed until the end of the study, it might appear that the treatment was effective merely because we were excluding individuals for whom the treatment was harmful. Thus, it is important to report the extent of missingness, whether missingness is systematically related to demographic and study variables, and to leverage modern approaches for handling missing data so that one does not exclude participants in the analysis merely because they have missing data (Nicholson et al., 2017). Multiple imputation and full information maximum likelihood are useful if data are missing at random or completely at random (Nicholson et al., 2017). Enders (2022, 2025) provides contemporary guidelines for handling missing data, including when data are missing not at random using selection models and pattern mixture models—including in longitudinal studies (Enders, 2022). Nicholson et al. (2017) provides recommendations for reporting about attrition and missingness in longitudinal studies.

Change in Meaning of the Scores

Another potential confound of change is if the same score has different meaning (with respect to the construct) at different points in development. That is, the scores are not on the same number line across time—they do not have the same scale or location across ages. If scores do not have the same scale or location across ages, any inferences about change (or lack thereof) could be incorrect. Consider, for example, if a behavior such as physical aggression becomes more severe (i.e., higher difficulty) with age. In this example, the same score—i.e., endorsement of physical aggression—would reflect a higher level of externalizing behavior in adulthood compared to early childhood. Or, consider if a behavior such as disobedience becomes less construct-valid (i.e., lower discrimination) with age. In this example, the same score difference—i.e., between endorsement (1) versus non-endorsement (0) of disobedience to authority—would reflect a greater difference in the level of externalizing behavior in early childhood compared to adulthood. Thus,

it is important to consider and account for changes in the meaning of scores across development. If a measure changes in the construct it assesses, this is called construct shift. Next, we describe longitudinal factorial invariance as a way to evaluate—and account for—the possibility that scores have changed in meaning across ages.

Factorial Invariance Across Time

Traditionally, it had been considered necessary to establish factorial invariance⁶ of the measure(s) across time—that is, the measure(s) assess the construct in the same way across time—to study individuals' change in those constructs ([Widaman et al., 2010](#)). As described earlier, a goal of longitudinal factorial invariance is to increase confidence that the latent factor at multiple timepoints is on the same scale so that meaningful changes in people's level can be observed. It had been considered necessary to establish at least strong factorial invariance—i.e., invariance of intercepts and factor loadings in factor analysis, or invariance of difficulty and discrimination in item response theory. Otherwise, a researcher may wrongly conclude that levels in the construct change over time when the apparent changes in the observed scores (and the latent variables) are driven by noninvariant loadings and/or intercepts for one or more items ([Lai, 2023](#)). However, it is now understood that violations of longitudinal factorial invariance do not mean that questions of change cannot be answered ([Lai, 2023](#)). Strong factorial invariance frequently does not hold, so in such situations, it may be necessary to find ways to accommodate or adjust for noninvariance ([Lai, 2023](#)).

At least partial strong factorial invariance (i.e., invariance of intercepts and factor loadings for a subset of items—the anchor items) is often considered a requirement for modeling individuals' change, such that the more items that show invariance, the greater the confidence that the latent factor is on a comparable scale at each age for examining individuals' change. However, even that requirement may be too strong. Researchers have argued that, in many cases, establishing factorial invariance may be unnecessary to compare latent means ([De Los Reyes et](#)

⁶ Factorial invariance is also called measurement invariance. In this context, we prefer the term “factorial invariance” because, as described, the actual measures used may differ across time.

al., 2025; Funder & Gardiner, 2024; Hancock & Curran, 2022; Knight & Zerr, 2010; Petersen et al., 2020; Robitzsch & Lüdtke, 2023; Welzel, Brunkert, et al., 2023; Welzel, Kruse, et al., 2023).

Indeed, failed longitudinal invariance can be an artifact of differing means across groups/time rather than bias or differences in the functioning of the measure (Welzel, Brunkert, et al., 2023; Welzel, Kruse, et al., 2023). As noted by Hancock (2022), to model differences in latent means (e.g., change over time), you do not need invariance; “you need to have a model that is correct...That means that...you are modeling [items’] invariance properly.”

When a construct demonstrates heterotypic continuity, longitudinal factorial invariance would not necessarily be expected because the construct’s factor structure, by definition, changes with development (Petersen et al., 2020). That is, noninvariances across development can be expected by theory (Stern et al., 2026). As a result, theoretical considerations are crucial and establishing even partial longitudinal factorial invariance may not be required when the construct shows heterotypic continuity (Knight & Zerr, 2010; Petersen et al., 2020). As noted by Millsap (2011, p. 1), “Measurement invariance is built on the notion that a measuring device should function in the same way across varied conditions, *so long as those varied conditions are irrelevant to the attribute being measured*” (emphasis added). However, age/development would not be considered irrelevant to psychopathology. As described earlier, externalizing (antisocial) behavior is considered to demonstrate heterotypic continuity; therefore, longitudinal factorial *noninvariance* for measures assessing the development of externalizing behavior may be expected.

As noted by De Los Reyes et al. (2025, p. 301), “developmental variations in antisocial behavior are unlikely to reflect confounds or biases in measurement.” De Los Reyes and colleagues characterize such tests of invariance across development—in which the conditions (age) are relevant to the construct of interest—as a questionable research practice, noting, “To be justified in testing measurement invariance [across ages], one would have to assume that when systematic differences manifest between ratings...for younger versus older youth, they cannot possibly reflect anything valid. For a scholar to make such an assumption, they must again ignore decades of theory and research in developmental psychopathology.” (p. 308). Moreover,

removing noninvariant items is not recommended if the item retains construct validity (De Los Reyes et al., 2025; Knight & Zerr, 2010). When the construct changes in manifestation, models with failed longitudinal factorial invariance can yield valid inferences. For instance, simulation work demonstrated that IRT linking approaches accurately captured changes in level on the latent factor even in the present of DIF and changing dimensionality (Edwards & Wirth, 2012).

There are multiple ways of handling noninvariance as described above, including alignment optimization, MNLFA, and IRT-based linking. In addition, another approach is to allow approximate factorial invariance. Approximate measurement invariance allows invariance parameters to have a small amount of noninvariance rather than fixing invariance parameters to zero (Van De Schoot et al., 2013). Approximate measurement invariance is more accurate than full or partial measurement invariance for estimating true latent mean differences when there are many small differences in the intercepts and factor loadings (Cieciuch et al., 2014; Van De Schoot et al., 2013). Approximate measurement invariance has also been applied in the longitudinal context (e.g., Seddig & Leitgöb, 2018; Winter & Depaoli, 2020). Robitzsch and Lüdtke (2023, p. 859) argue that establishing factorial invariance is neither necessary nor sufficient for comparing latent means, and they note that “there is no support for preferring the partial invariance approach over competing approaches, such as invariance alignment, robust linking, or Bayesian approximate invariance.”

Integration Across Multiple Sources

The most accurate informant(s) for a person can change with development, so it may be necessary to include different informants across time. There is widespread agreement that including multiple sources of data—e.g., multiple informants or multiple measurement methods—is preferable. However, there is less agreement about how to use multiple sources of data (Bauer et al., 2013), particularly because informants often show low correspondence with one another when rating the same person. Some of these discrepancies may be confounded by the person behaving differently in different contexts or with different people. The inclusion of multiple measurement methods can help account for measurement error given the considerable

variance that is attributable to method bias ([Podsakoff et al., 2012](#)). The inclusion of multiple informants can be relevant for informing about the pervasiveness of a person's behavior across contexts (e.g., home, school).

When including multiple or different informants across development, it is important to consider how to leverage the multi-informant ratings. Some researchers advocate for integrating scores using principal component approaches (while disentangling variance attributable to trait, perspective, and context; [Kraemer et al., 2003](#)) or latent variable approaches (e.g., [Bauer et al., 2013](#)), and others advocate for keeping different informants' ratings separate without integrating them (but including them in the same model; [Aitken et al., 2023](#)). As an example of a latent variable modeling approach, Bauer et al. (2013) demonstrate a trifactor model that attempts to disentangle construct variance from the method effects tied to the specific item and the informant. De Los Reyes et al. (2013) provide the Operations Triad Model as a framework for using and interpreting reports from multiple informants (for more information, see [Makol et al., 2025](#)). In sum, finding ways to leverage the information from multiple informants will be crucial for studying the development of psychopathology across the lifespan ([De Los Reyes, 2026](#)).

Obsolescence of Items and Measures

Another challenge is that questionnaire items—whether the particular wording or the behavior itself—can become obsolete and irrelevant with secular changes. For instance, due to technological changes, items focused on television use—without considering other forms of screen use—may miss important contemporary screen-related use. Additionally, ways people use substances change over historical time. For instance, vaping is a more contemporary way that many adolescents use nicotine. The ways in which aggressive behavior manifests also change across historical time. Items focused on physical aggression would miss important contemporary manifestations of aggression, including cyberaggression, cyberbullying, online harassment, and aggression on social media (e.g., exclusion). In sum, it is important to keep items relevant for current use, and that naturally means changing item content and/or wording across historical time, a further reason linking of scores (given differing item content) is crucial.

Conclusion

Due to a variety of challenges—both conceptual and methodological—that we discuss, the field has not adequately studied individuals' development of psychopathology across the lifespan, instead taking a piecewise approach to development that limits our understanding of lifespan development and the origins of psychopathology. To address these challenges, we discuss the importance of defining psychopathology constructs from a developmental perspective, developing new assessments that (a) capture changes in behavioral manifestation of psychopathology, (b) apply scoring systems that allow charting individuals' change over time in the face of changing behavioral manifestation, and (c) capture the full range of individual differences. We introduce a developmental scaling framework to help address these challenges. Addressing these challenges will allow charting individuals' trajectories of psychopathology across the lifespan and identifying risk and protective factors that influence their developmental courses. Doing so is critical for understanding how psychopathology develops, when to intervene, and how to most effectively prevent psychopathology. Although this paper focuses on development of psychopathology, the principles discussed likely apply to many other developmental domains.

References

- Abrous, D. N., Blin, N., Boraxbekk, C.-J., Catheline, G., Fitzsimons, C. P., Hilscher, M., Lemoine, M., Lopes, L. V., Maass, A., Nilsson, M., Nyberg, L., Rasmussen, L. J., Remondes, M., Sauvage, M., Schreiber, S., & Wolbers, T. (2026). Hallmarks of healthy cognitive aging: Inter-individual differences in aging trajectories. *Ageing Research Reviews*, 119, 103102. <https://doi.org/10.1016/j.arr.2026.103102>
- Ackerman, P. L. (2013). Assessment of intellectual functioning in adults. In K. F. Geisinger, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology: 2: Testing and assessment in clinical and counseling psychology* (pp. 119–132). American Psychological Association. <https://doi.org/10.1037/14048-008>
- Aiken, A., Clare, P. J., Wadolowski, M., Hutchinson, D., Najman, J. M., Slade, T., Bruno, R.,

- McBride, N., Kypri, K., & Mattick, R. P. (2018). Age of alcohol initiation and progression to binge drinking in adolescence: A prospective cohort study. *Alcoholism: Clinical and Experimental Research*, 42(1), 100–110. <https://doi.org/10.1111/acer.13525>
- Aitken, M., Plamondon, A., Krzeczkowski, J., Kil, H., & Andrade, B. F. (2023). Systematic integration of multi-informant externalizing ratings in clinical settings. *Research on Child and Adolescent Psychopathology*. <https://doi.org/10.1007/s10802-023-01119-z>
- American Psychiatric Association. (2022). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR)*. American Psychiatric Publishing.
- Applegate, B., & Lahey, B. B. (2026). Mapping persistence and change in psychological problems during the transition to adolescence: Adding, subtracting, shifting, and persisting. *JCPP Advances*, 6(2), e70043. <https://doi.org/10.1002/jcv2.70043>
- Averdijk, M. (2014). Methodological challenges in the study of age–victimization patterns. Can we use the accelerated design of the NCVS to reconstruct victim careers? *International Review of Victimology*, 20(3), 265–288. <https://doi.org/10.1177/0269758014537147>
- Barry, T. D., Marcus, D. K., Barry, C. T., & Coccaro, E. F. (2013). The latent structure of oppositional defiant disorder in children and adults. *Journal of Psychiatric Research*, 47(12), 1932–1939. <https://doi.org/10.1016/j.jpsychires.2013.08.016>
- Bates, J. E., Schermerhorn, A. C., & Petersen, I. T. (2014). Temperament concepts in developmental psychopathology. In K. Rudolph & M. Lewis (Eds.), *Handbook of developmental psychopathology* (3rd ed., pp. 311–329). Springer. https://doi.org/10.1007/978-1-4614-9608-3_16
- Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, 22(3), 507–526. <https://doi.org/10.1037/met0000077>
- Bauer, D. J., Howard, A. L., Baldasaro, R. E., Curran, P. J., Hussong, A. M., Chassin, L., & Zucker, R. A. (2013). A trifactor model for integrating ratings across multiple informants. *Psychological Methods*, 18(4), 475–493. <https://doi.org/10.1037/a0032475>

- Berluti, K., Amormino, P., Potter, A., Wshah, S., & Marsh, A. (2026). Machine learning prediction of conduct problems in children using the longitudinal ABCD Study. *Journal of Child Psychology and Psychiatry*, 67(3), 390–399. <https://doi.org/10.1111/jcpp.70057>
- Bista, S., Graham, P. L., Kang, M., Lin, A., Lymer, S., Marino, J. L., Robinson, M., Skinner, S. R., Steinbeck, K., Straker, L. M., & Tait, R. J. (2025). Joint developmental trajectories of internalizing and externalizing problems from mid-childhood to late adolescence and childhood risk factors: Findings from a prospective pre-birth cohort. *Development and Psychopathology*, 37(1), 176–191. <https://doi.org/10.1017/S0954579423001505>
- Bolhuis, K., Lubke, G. H., Ende, J. van der, Bartels, M., Beijsterveldt, C. E. M. van, Lichtenstein, P., Larsson, H., Jaddoe, V. W. V., Kushner, S. A., Verhulst, F. C., Boomsma, D. I., & Tiemeier, H. (2017). Disentangling heterogeneity of childhood disruptive behavior problems into dimensions and subgroups. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(8), 678–686. <https://doi.org/10.1016/j.jaac.2017.05.019>
- Bongers, I. L., Koot, H. M., Ende, J. van der, & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*, 75(5), 1523–1537. <https://doi.org/10.1111/j.1467-8624.2004.00755.x>
- Borelli, W. V., Carmona, K. C., Studart-Neto, A., Nitrini, R., Caramelli, P., & Costa, J. C. da. (2018). Operationalized definition of older adults with high cognitive performance. *Dementia & Neuropsychologia*, 12(3), 221–227. <https://doi.org/10.1590/1980-57642018dn12-030001>
- Briggs-Gowan, M. J., Godoy, L., Heberle, A., & Carter, A. S. (2016). Assessment of psychopathology in young children. In D. Cicchetti (Ed.), *Developmental psychopathology: 1: Theory and method* (pp. 1–45). <https://doi.org/10.1002/9781119125556.devpsy101>
- Broeren, S., Muris, P., Diamantopoulou, S., & Baker, J. R. (2013). The course of childhood anxiety symptoms: Developmental trajectories and child-related factors in normal children. *Journal of Abnormal Child Psychology*, 41(1), 81–95.

<https://doi.org/10.1007/s10802-012-9669-9>

- Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., Fergusson, D., Horwood, J. L., Loeber, R., Laird, R., Lynam, D. R., Moffitt, T. E., Pettit, G. S., & Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology*, 39(2), 222–245. <https://doi.org/10.1037/0012-1649.39.2.222>
- Burns, G. L., Walsh, J. A., Patterson, D. R., Holte, C. S., Sommers-Flanagan, R., & Parker, C. M. (2001). Attention deficit and disruptive behavior disorder symptoms: Usefulness of a frequency count rating procedure to measure these symptoms. *European Journal of Psychological Assessment*, 17(1), 25–35. <https://doi.org/10.1027/1015-5759.17.1.25>
- Burt, S. A. (2012). How do we optimally conceptualize the heterogeneity within antisocial behavior? An argument for aggressive versus non-aggressive behavioral dimensions. *Clinical Psychology Review*, 32(4), 263–279. <https://doi.org/10.1016/j.cpr.2012.02.006>
- Caspi, A., Houts, R. M., Ambler, A., Danese, A., Elliott, M. L., Hariri, A., Harrington, H., Hogan, S., Poulton, R., Ramrakha, S., Rasmussen, L. J. H., Reuben, A., Richmond-Rakerd, L., Sugden, K., Wertz, J., Williams, B. S., & Moffitt, T. E. (2020). Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the Dunedin birth cohort study. *JAMA Network Open*, 3(4), e203221–e203221. <https://doi.org/10.1001/jamanetworkopen.2020.3221>
- Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, 2(2), 119–137. <https://doi.org/10.1177/2167702613497473>
- Caspi, A., & Shiner, R. L. (2006). Personality development. In N. Eisenberg, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (6th ed., Vol. 3, pp. 300–365). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470147658.chpsy0306>
- Chen, F. R., & Jaffee, S. R. (2015). The heterogeneity in the development of homotypic and

- heterotypic antisocial behavior. *Journal of Developmental and Life-Course Criminology*, 1(3), 269–288. <https://doi.org/10.1007/s40865-015-0012-3>
- Chen, S. M., & Bauer, D. J. (in press a). Improving the evaluation of construct change over time: Advantages of longitudinal moderated nonlinear factor analysis over conventional first-order growth models. *Multivariate Behavioral Research*, 1–27. <https://doi.org/10.1080/00273171.2026.2640576>
- Chen, S. M., & Bauer, D. J. (in press b). Modeling construct change over time amidst potential changes in construct measurement: A longitudinal moderated factor analysis approach. *Psychological Methods*. <https://doi.org/10.1037/met0000685>
- Chumlea, W. M. C., Choh, A., Lee, M., Towne, B., Duren, D., Siervogel, R. M., & Czerwinski, S. (2012). Maintaining function with aging what we have learned from the Fels Longitudinal Study. *The Journal of Frailty & Aging*, 1(2), 50–51. <https://doi.org/10.14283/jfa.2012.9>
- Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting and Clinical Psychology*, 70(1), 6–20. <https://doi.org/10.1037/0022-006X.70.1.6>
- Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, 5(982). <https://doi.org/10.3389/fpsyg.2014.00982>
- Coghill, D., & Sonuga-Barke, E. J. S. (2012). Annual Research Review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders: Implications of recent empirical study. *Journal of Child Psychology and Psychiatry*, 53(5), 469–489. <https://doi.org/10.1111/j.1469-7610.2011.02511.x>
- Culpepper, S. A. (2013). The reliability and precision of total scores and IRT estimates as a function of polytomous IRT parameters and latent trait distribution. *Applied Psychological Measurement*, 37(3), 201–225. <https://doi.org/10.1177/0146621612470210>
- Curran, P. J., Hussong, A. M., Cai, L., Huang, W., Chassin, L., Sher, K. J., & Zucker, R. A. (2008).

- Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*, 44(2), 365.
<https://doi.org/10.1037/0012-1649.44.2.365>
- Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., & Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research*, 49(3), 214–231. <https://doi.org/10.1080/00273171.2014.889594>
- De Los Reyes, A. (2026). Toward integrative, multisource, and multimodal approaches to assessing psychopathology intergenerationally and across the lifespan: Commentary on Caspi et al. (2026). *Journal of Psychopathology and Clinical Science*, 135(4), 495–497.
<https://doi.org/10.1037/abn0001084>
- De Los Reyes, A., L., O. F., J., R. S., Armando, P., D., M. B., Mo, W., & Charamut, N. R. and. (2025). Editorial: Fairness, validity, and transparency in what researchers assume when testing for measurement invariance. *Journal of Clinical Child & Adolescent Psychology*, 54(3), 299–317. <https://doi.org/10.1080/15374416.2025.2484813>
- De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kunder, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, 9(1), 123–149. <https://doi.org/10.1146/annurev-clinpsy-050212-185617>
- Dodge, K. A., Bates, J. E., & Pettit, G. S. (1990). Mechanisms in the cycle of violence. *Science*, 250(4988), 1678–1683. <https://doi.org/10.1126/science.2270481>
- Edwards, M. C., & Wirth, R. J. (2012). Valid measurement without factorial invariance: A longitudinal example. In G. R. Hancock & J. R. Harring (Eds.), *Advances in longitudinal methods in the social and behavioral sciences* (pp. 289–311). Information Age Publishing.
<https://doi.org/10.1108/978-1-61735-891-320251015>
- Elliott, D. S., & Ageton, S. S. (1980). Reconciling race and class differences in self-reported and official estimates of delinquency. *American Sociological Review*, 45(1), 95–110.
<https://doi.org/10.2307/2095245>

- Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.
- Enders, C. K. (2025). Missing data: An update on the state of the art. *Psychological Methods*, 30(2), 322–339. <https://doi.org/10.1037/met0000563>
- Fairchild, G., Goozen, S. H. M. van, Calder, A. J., & Goodyer, I. M. (2013). Research Review: Evaluating and reformulating the developmental taxonomic theory of antisocial behaviour. *Journal of Child Psychology and Psychiatry*, 54(9), 924–940. <https://doi.org/10.1111/jcpp.12102>
- Farrington, D. P., Piquero, A. R., Jennings, W. G., & Jolliffe, D. (2023). *Offending from childhood to late middle age: Recent results from the Cambridge Study in Delinquent Development* (pp. 47–71). Springer New York. https://doi.org/10.1007/978-1-0716-3335-9_5
- Ferrer, E., Salthouse, T. A., Stewart, W. F., & Schwartz, B. S. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. *Psychology and Aging*, 19(2), 243–259. <https://doi.org/10.1037/0882-7974.19.2.243>
- Forbes, M. K., Tackett, J. L., Markon, K. E., & Krueger, R. F. (2016). Beyond comorbidity: Toward a dimensional and hierarchical approach to understanding psychopathology across the life span. *Development and Psychopathology*, 28(4pt1), 971–986. <https://doi.org/10.1017/S0954579416000651>
- Frick, P. J., & Morris, A. S. (2004). Temperament and developmental pathways to conduct problems. *Journal of Clinical Child & Adolescent Psychology*, 33(1), 54–68. https://doi.org/10.1207/S15374424JCCP3301_6
- Funder, D. C., & Gardiner, G. (2024). Misgivings about measurement invariance. *European Journal of Personality*, 38(6), 889–895. <https://doi.org/10.1177/08902070241228338>
- Glueck, S., & Glueck, E. (1950). *Unraveling juvenile delinquency*. Harvard University Press.
- Godoy, L. L. de, Alves, C. A. P. F., Saavedra, J. S. M., Studart-Neto, A., Nitrini, R., Costa Leite, C. da, & Bisdas, S. (2021). Understanding brain resilience in superagers: A systematic review. *Neuroradiology*, 63(5), 663–683. <https://doi.org/10.1007/s00234-020-02562-1>
- Gornik, A. E., Clark, D. A., Durbin, C. E., & Zucker, R. A. (2023). Individual differences in the

- development of youth externalizing problems predict a broad range of adult psychosocial outcomes. *Development and Psychopathology*, 35(2), 630–651.
<https://doi.org/10.1017/S0954579421001772>
- Grant, S. J., Aitchison, T., Henderson, E., Christie, J., Zare, S., McMurray, J., & Dargie, H. (1999). A comparison of the reproducibility and the sensitivity to change of visual analogue scales, Borg scales, and Likert scales in normal subjects during submaximal exercise. *Chest*, 116(5), 1208–1217. <https://doi.org/10.1378/chest.116.5.1208>
- Hancock, G. R., & Buehl, M. M. (2008). Second-order latent growth models with shifting indicators. *Journal of Modern Applied Statistical Methods*, 7(1), 39–55.
<https://doi.org/10.22237/jmasm/1209614640>
- Hancock, G. R., & Curran, P. J. (Hosts). (2022, February 1). *Quantitude: S3E18: Aunt Roz's guide to latent means models* [Audio recording]. In Quantitude; Quantitude.
<https://quantitudepod.org/latent-means/> (archived at: <https://perma.cc/5JL3-YUPA>)
- Harris, J. L., LeBeau, B., & Petersen, I. T. (2025). Reactive and control processes in the development of internalizing and externalizing problems across early childhood to adolescence. *Development and Psychopathology*, 37(2), 836–858.
<https://doi.org/10.1017/S0954579424000713>
- Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine*, 42(05), 903–920. <https://doi.org/10.1017/S0033291711001966>
- Haslam, N., McGrath, M. J., Viechtbauer, W., & Kuppens, P. (2020). Dimensions over categories: A meta-analysis of taxometric research. *Psychological Medicine*, 50(9), 1418–1432.
<https://doi.org/10.1017/S003329172000183X>
- Hertzog, C., & Nesselroade, J. R. (2003). Assessing psychological change in adulthood: An overview of methodological issues. *Psychology and Aging*, 18(4), 639–657.
<https://doi.org/10.1037/0882-7974.18.4.639>
- Hosch, A., Oleson, J. J., Harris, J. L., Goeltz, M. T., Neumann, T., LeBeau, B., Hazeltine, E., &

- Petersen, I. T. (2022). Studying children's growth in self-regulation using changing measures to account for heterotypic continuity: A Bayesian approach to developmental scaling. *Developmental Science*, 25(6), e13280. <https://doi.org/10.1111/desc.13280>
- Insel, T. R. (2008). Assessing the economic costs of serious mental illness. *American Journal of Psychiatry*, 165(6), 663–665. <https://doi.org/10.1176/appi.ajp.2008.08030366>
- Kaat, A. J., Blackwell, C. K., Estabrook, R., Burns, J. L., Petittclerc, A., Briggs-Gowan, M. J., Gershon, R. C., Cella, D., Perlman, S. B., & Wakschlag, L. S. (2019). Linking the Child Behavior Checklist (CBCL) with the Multidimensional Assessment Profile of Disruptive Behavior (MAP-DB): Advancing a dimensional spectrum approach to disruptive behavior. *Journal of Child and Family Studies*, 28(2), 343–353. <https://doi.org/10.1007/s10826-018-1272-4>
- Kagan, J. (1969). The three faces of continuity in human development. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 983–1002). Rand McNally.
- Kagan, J. (1971). *Change and continuity in infancy*. Wiley.
- Kagan, J. (1980). Perspectives on continuity. In O. G. Brim Jr. & J. Kagan (Eds.), *Constancy and change in human development* (pp. 26–74). Harvard University Press.
- Kagan, J., & Moss, H. A. (1962). *Birth to maturity: A study in psychological development*. Wiley. <https://doi.org/10.1037/13129-000>
- Keijsers, L., Loeber, R., Branje, S., & Meeus, W. (2012). Parent-child relationships of boys in different offending trajectories. A developmental perspective. *Journal of Child Psychology and Psychiatry*, 53(12), 1222–1232. <https://doi.org/10.1111/j.1469-7610.2012.02585.x>
- Kim, H. K., Capaldi, D. M., Pears, K. C., Kerr, D. C. R., & Owen, L. D. (2009). Intergenerational transmission of internalising and externalising behaviours across three generations: Gender-specific pathways. *Criminal Behaviour and Mental Health*, 19(2), 125–141. <https://doi.org/10.1002/cbm.708>
- Kjeldsen, A., Nilsen, W., Gustavson, K., Skipstein, A., Melkevik, O., & Karevold, E. B. (2016). Predicting well-being and internalizing symptoms in late adolescence from trajectories of

- externalizing behavior starting in infancy. *Journal of Research on Adolescence*, 26(4), 991–1008. <https://doi.org/10.1111/jora.12252>
- Kliem, S., Heinrichs, N., Lohmann, A., Bussing, R., Schwarzer, G., & Briegel, W. (2018). Dimensional latent structure of early disruptive behavior disorders: A taxometric analysis in preschoolers. *Journal of Abnormal Child Psychology*, 46, 1385–1394. <https://doi.org/10.1007/s10802-017-0383-5>
- Kliem, S., Krieg, Y., Klatt, T., & Baier, D. (2022). Dimensional latent structure of callous-unemotional traits in german adolescents: Results from taxometric analyses. *Research on Child and Adolescent Psychopathology*, 50(6), 771–780. <https://doi.org/10.1007/s10802-021-00885-y>
- Knight, G. P., & Zerr, A. A. (2010). Informed theory and measurement equivalence in child development research. *Child Development Perspectives*, 4(1), 25–30. <https://doi.org/10.1111/j.1750-8606.2009.00112.x>
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed., p. 566). Springer. <https://doi.org/10.1007/978-1-4939-0317-7>
- Korhonen, M., Luoma, I., Salmelin, R., Siirtola, A., & Puura, K. (2018). The trajectories of internalizing and externalizing problems from early childhood to adolescence and young adult outcome. *Journal of Child and Adolescent Psychiatry*, 2(3), 7–12.
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454–477. <https://doi.org/10.1037/abn0000258>
- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of*

- Psychiatry*, 160(9), 1566–1577. <https://doi.org/10.1176/appi.ajp.160.9.1566>
- Krueger, R. F., & DeYoung, C. G. (2016). The RDoC initiative and the structure of psychopathology. *Psychophysiology*, 53(3), 351–354. <https://doi.org/10.1111/psyp.12551>
- Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Keyes, K. M., Latzman, R. D., Michelini, G., Patrick, C. J., Sellbom, M., Slade, T., South, S. C., Sunderland, M., Tackett, J. L., Waldman, I., Waszczuk, M. A., ... Workgroup, H. U. (2021). Validity and utility of hierarchical taxonomy of psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry*, 20(2), 171–193. <https://doi.org/10.1002/wps.20844>
- Krueger, R. F., & Markon, K. E. (2011). A dimensional-spectrum model of psychopathology: Progress and opportunities. *Archives of General Psychiatry*, 68(1), 10–11. <https://doi.org/10.1001/archgenpsychiatry.2010.188>
- Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology*, 116(4), 645–666. <https://doi.org/10.1037/0021-843X.116.4.645>
- Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, 114(4), 537–550. <https://doi.org/10.1037/0021-843x.114.4.537>
- Krueger, R. F., Nichol, P. E., Hicks, B. M., Markon, K. E., Patrick, C. J., Iacono, W. G., & McGue, M. (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment*, 16(2), 107–119. <https://doi.org/10.1037/1040-3590.16.2.107>
- Krueger, R. F., & Piasecki, T. M. (2002). Toward a dimensional and psychometrically-informed approach to conceptualizing psychopathology. *Behaviour Research and Therapy*, 40(5), 485–499. [https://doi.org/10.1016/S0005-7967\(02\)00016-5](https://doi.org/10.1016/S0005-7967(02)00016-5)
- Lai, M. H. C. (2023). Adjusting for measurement noninvariance with alignment in growth

- modeling. *Multivariate Behavioral Research*, 58(1), 30–47.
<https://doi.org/10.1080/00273171.2021.1941730>
- Lansford, J. E., Malone, P. S., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2010). Developmental cascades of peer rejection, social information processing biases, and aggression during middle childhood. *Development and Psychopathology*, 22(3), 593–602.
<https://doi.org/10.1017/S0954579410000301>
- Le Blanc, M. (2021). *The development of antisocial behavior and crime*. Springer.
<https://doi.org/10.1007/978-3-030-68429-7>
- Little, T. D. (2024). *Longitudinal structural equation modeling* (2nd ed.). Guilford Publications.
- Makol, B. A., Yang, J., Wang, M., & De Los Reyes, A. (2025). *Youth development in context: Integrating multiple informants to assess behavior*. Springer Nature Switzerland.
<https://doi.org/10.1007/978-3-031-80549-3>
- Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*, 137(5), 856–879. <https://doi.org/10.1037/a0023678>
- Markon, K. E., & Krueger, R. F. (2005). Categorical and continuous models of liability to externalizing disorders: A direct comparison in NESARC. *Archives of General Psychiatry*, 62(12), 1352–1359. <https://doi.org/10.1001/archpsyc.62.12.1352>
- Masche, J. G., & Dulmen, M. H. M. van. (2004). Advances in disentangling age, cohort, and time effects: No quadrature of the circle, but a help. *Developmental Review*, 24(3), 322–342.
<https://doi.org/10.1016/j.dr.2004.04.002>
- McArdle, J. J., & Grimm, K. J. (2011). An empirical example of change analysis by linking longitudinal item response data from multiple tests. In A. A. von Davier (Ed.), *Statistical models for test equating, scaling, and linking* (pp. 71–88). Springer Science & Business Media.
- McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and

- changing scales of measurement. *Psychological Methods*, 14(2), 126–149.
<https://doi.org/10.1037/a0015857>
- McDaniel, H. L., Saavedra, L. M., Morgan-López, A. A., Bradshaw, C. P., Lochman, J. E., Kaihoi, C. A., Powell, N. P., Qu, L., & Yaros, A. C. (2023). Harmonizing social, emotional, and behavioral constructs in prevention science: Digging into the weeds of aligning disparate measures. *Prevention Science*. <https://doi.org/10.1007/s11121-022-01467-0>
- McDermott, P. A., Watkins, M. W., Rovine, M. J., & Rikoon, S. H. (2013). Assessing changes in socioemotional adjustment across early school transitions New national scales for children at risk. *Journal of School Psychology*, 51(1), 97–115.
<https://doi.org/10.1016/j.jsp.2012.10.002>
- Miller, J. L., Vaillancourt, T., & Boyle, M. H. (2009). Examining the heterotypic continuity of aggression using teacher reports: Results from a national Canadian study. *Social Development*, 18(1), 164–180. <https://doi.org/10.1111/j.1467-9507.2008.00480.x>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Taylor & Francis.
- Mineo, L. (2017). *Good genes are nice, but joy is better*.
<https://news.harvard.edu/gazette/story/2017/04/over-nearly-80-years-harvard-study-has-been-showing-how-to-live-a-healthy-and-happy-life/>
- Moeller, J. (2015). A word on standardization in longitudinal studies: Don't. *Frontiers in Psychology*, 6(1389), 1–4. <https://doi.org/10.3389/fpsyg.2015.01389>
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100(4), 674–701.
<https://doi.org/10.1037/0033-295X.100.4.674>
- Moffitt, T. E., Arseneault, L., Belsky, D. W., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698.
<https://doi.org/10.1073/pnas.1010076108>

- Moss, H. A., & Kagan, J. (1964). Report on personality consistency and change from the Fels Longitudinal Study. *Vita Humana*, 7(2), 127–138. <https://doi.org/10.1159/000270073>
- Murayama, K., Pekrun, R., Lichtenfeld, S., & Hofe, R. vom. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development*, 84(4), 1475–1490. <https://doi.org/10.1111/cdev.12036>
- Muthén, B. O., & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in Psychology*, 5(978). <https://doi.org/10.3389/fpsyg.2014.00978>
- Nicholson, J. S., Deboeck, P. R., & Howard, W. (2017). Attrition in developmental psychology: A review of modern missing data reporting and practices. *International Journal of Behavioral Development*, 41(1), 143–153. <https://doi.org/10.1177/0165025415618275>
- Nickerson, A. B., & Fredrick, S. S. (2026). Cyberbullying and mental health in children and adolescents. *Annual Review of Clinical Psychology*, 22, 559–586. <https://doi.org/10.1146/annurev-clinpsy-061324-073435>
- Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, 20(2), 673–716. <https://doi.org/10.1017/S0954579408000333>
- Oleson, J. J., Cavanaugh, J. E., Tomblin, J. B., Walker, E., & Dunn, C. (2016). Combining growth curves when a longitudinal study switches measurement tools. *Statistical Methods in Medical Research*, 25(6), 2925–2938. <https://doi.org/10.1177/0962280214534588>
- Owens, E. B., & Shaw, D. S. (2003). Predicting growth curves of externalizing behavior across the preschool years. *Journal of Abnormal Child Psychology*, 31(6), 575–590. <https://doi.org/10.1023/a:1026254005632>
- Paalman, C. H., Terwee, C. B., Jansma, E. P., & Jansen, L. M. C. (2013). Instruments measuring externalizing mental health problems in immigrant ethnic minority youths: A systematic review of measurement properties. *PLOS ONE*, 8(5), e63109.

<https://doi.org/10.1371/journal.pone.0063109>

Patterson, G. R. (1993). Orderly change in a stable world: The antisocial trait as a chimera.

Journal of Consulting and Clinical Psychology, 61(6), 911–919.

<https://doi.org/10.1037/0022-006X.61.6.911>

Pavlovich, K., Tiego, J., Constable, T., Fornito, A., & Imaging, M. B. (2026). Assessing the psychometric properties of the Child Behavior Checklist in the ABCD Study.

Developmental Science, 29(4), e70216. <https://doi.org/10.1111/desc.70216>

Petersen, I. T. (2024). Reexamining developmental continuity and discontinuity in the 21st century: Better aligning behaviors, functions, and mechanisms. *Developmental*

Psychology, 60(11), 1992–2007. <https://doi.org/10.1037/dev0001657>

Petersen, I. T. (2026). *Principles of psychological assessment: With applied examples in R*.

University of Iowa Libraries. <https://doi.org/10.25820/work.007199>

Petersen, I. T., Bates, J. E., Dodge, K. A., Lansford, J. E., & Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood.

Development and Psychopathology, 27(3), 791–818.

<https://doi.org/10.1017/S0954579414000789>

Petersen, I. T., Choe, D. E., & LeBeau, B. (2020). Studying a moving target in development: The challenge and opportunity of heterotypic continuity. *Developmental Review*, 58, 100935.

<https://doi.org/10.1016/j.dr.2020.100935>

Petersen, I. T., Demko, Z., Doebler, P., Sabel, L., Oleson, J. J., & Krueger, R. F. (2026). How often is “often”? Improving assessment of the externalizing spectrum using absolute frequency.

Psychological Assessment, 38(4), 295–306. <https://doi.org/10.1037/pas0001441>

Petersen, I. T., Demko, Z., Lee, W.-C., & Oleson, J. J. (2026). Studying development of

psychopathology using changing measures to account for heterotypic continuity. *JAACAP*

Open, 4(1), 111–123. <https://doi.org/10.1016/j.jaacop.2025.10.008>

Petersen, I. T., & LeBeau, B. (2021). Language ability in the development of externalizing

behavior problems in childhood. *Journal of Educational Psychology*, 113(1), 68–85.

<https://doi.org/10.1037/edu0000461>

Petersen, I. T., & LeBeau, B. (2022). Creating a developmental scale to chart the development of psychopathology with different informants and measures across time. *Journal of Psychopathology and Clinical Science*, 131(6), 611–625.

<https://doi.org/10.1037/abn0000649>

Petersen, I. T., LeBeau, B., & Choe, D. E. (2021). Creating a developmental scale to account for heterotypic continuity in development: A simulation study. *Child Development*, 92(1), e1–e19. <https://doi.org/10.1111/cdev.13433>

Petersen, I. T., Lindhiem, O., LeBeau, B., Bates, J. E., Pettit, G. S., Lansford, J. E., & Dodge, K. A. (2018). Development of internalizing problems from adolescence to emerging adulthood: Accounting for heterotypic continuity with vertical scaling. *Developmental Psychology*, 54(3), 586–599. <https://doi.org/10.1037/dev0000449>

Petscher, Y., Justice, L. M., & Hogan, T. (2018). Modeling the early language trajectory of language development when the measures change and its relation to poor reading comprehension. *Child Development*, 89(6), 2136–2156.

<https://doi.org/10.1111/cdev.12880>

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63(1), 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>

Poulton, R., Moffitt, T. E., & Silva, P. A. (2015). The Dunedin Multidisciplinary Health and Development Study: Overview of the first 40 years, with an eye to the future. *Social Psychiatry and Psychiatric Epidemiology*, 50(5), 679–693.

<https://doi.org/10.1007/s00127-015-1048-8>

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>

Robitzsch, A., & Lüdtke, O. (2023). Why full, partial, or approximate measurement invariance are not a prerequisite for meaningful and valid group comparisons. *Structural Equation*

- Modeling: A Multidisciplinary Journal*, 30(6), 859–870.
<https://doi.org/10.1080/10705511.2023.2191292>
- Roche, A. F. (1992). *Growth, maturation, and body composition: The Fels Longitudinal Study 1929–1991*. Cambridge University Press.
- Rohrer, J. M. (2025). Thinking clearly about age, period, and cohort effects. *Advances in Methods and Practices in Psychological Science*, 8(2), 25152459251342750.
<https://doi.org/10.1177/25152459251342750>
- Rutter, M., Kim-Cohen, J., & Maughan, B. (2006). Continuities and discontinuities in psychopathology between childhood and adult life. *Journal of Child Psychology and Psychiatry*, 47(3-4), 276–295. <https://doi.org/10.1111/j.1469-7610.2006.01614.x>
- Sacu, S., Streit, F., Witt, S. H., Banaschewski, T., Holz, N. E., & Schultebrasucks, K. (in press). Predicting psychopathology symptom trajectories using machine learning: A 33-year prospective study. *Journal of Child Psychology and Psychiatry*.
<https://doi.org/10.1111/jcpp.70180>
- Schaefer, J. D., Caspi, A., Belsky, D. W., Harrington, H., Houts, R., Horwood, L. J., Hussong, A., Ramrakha, S., Poulton, R., & Moffitt, T. E. (2017). Enduring mental health: Prevalence and prediction. *Journal of Abnormal Psychology*, 126(2), 212–224.
<https://doi.org/10.1037/abn0000232>
- Schaeffer, N. C. (1991). Hardly ever or constantly? Group comparisons using vague quantifiers. *Public Opinion Quarterly*, 55(3), 395–423. <https://doi.org/10.1086/269270>
- Schaie, K. W. (2005). *Developmental influences on adult intelligence: The Seattle Longitudinal Study*. Oxford University Press.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93–105. <https://doi.org/10.1037/0003-066X.54.2.93>
- Seddig, D., & Leitgöb, H. (2018). Approximate measurement invariance and longitudinal confirmatory factor analysis: Concept and application with panel data. *Survey Research Methods*, 12(1). <https://doi.org/10.18148/srm/2018.v12i1.7210>

- Sellbom, M. (2016). Elucidating the validity of the externalizing spectrum of psychopathology in correctional, forensic, and community samples. *Journal of Abnormal Psychology, 125*(8), 1027–1038. <https://doi.org/10.1037/abn0000171>
- Shi, Q., Ettetal, I., Deutz, M. H. F., & Woltering, S. (2020). Trajectories of pure and co-occurring internalizing and externalizing problems from early childhood to adolescence: Associations with early childhood individual and contextual antecedents. *Developmental Psychology, 56*(10), 1906–1918. <https://doi.org/10.1037/dev0001095>
- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment, 31*(4), 557–566. <https://doi.org/10.1037/pas0000648>
- Sivertsson, F., Carlsson, C., Almquist, Y. B., & Brännström, L. (2024). Offending trajectories from childhood to retirement age: Findings from the Stockholm birth cohort study. *Journal of Criminal Justice, 91*, 102155. <https://doi.org/10.1016/j.jcrimjus.2024.102155>
- Speyer, L. G., Neaves, S., Hall, H. A., Hemani, G., Lombardo, M. V., Murray, A. L., Auyeung, B., & Luciano, M. (2022). Polygenic risks for joint developmental trajectories of internalizing and externalizing problems: Findings from the ALSPAC cohort. *Journal of Child Psychology and Psychiatry, 63*(8), 948–956. <https://doi.org/10.1111/jcpp.13549>
- Sroufe, L. A. (2016). The place of attachment in development. In J. Cassidy & P. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (Vol. 3, pp. 997–1011). Guilford.
- Steffensmeier, D., Slepicka, J., & Schwartz, J. (2025). International and historical variation in the age–crime curve. *Annual Review of Criminology, 8*, 239–268. <https://doi.org/10.1146/annurev-criminol-111523-122451>
- Sterba, S. K., Prinstein, M. J., & Cox, M. J. (2007). Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences. *Development and Psychopathology, 19*(2), 345–366. <https://doi.org/10.1017/S0954579407070174>
- Sterner, P., Pargent, F., & Goretzko, D. (2026). Don't let MI be misunderstood: A causal view on

- measurement invariance. *Current Research in Ecological and Social Psychology*, 10, 100261. <https://doi.org/10.1016/j.cresp.2025.100261>
- Stringer, D., Kent, R., Briskman, J., Lukito, S., Charman, T., Baird, G., Lord, C., Pickles, A., & Simonoff, E. (2020). Trajectories of emotional and behavioral problems from childhood to early adult life. *Autism*, 24(4), 1011–1024. <https://doi.org/10.1177/1362361320908972>
- Student, S. R. (2025). Vertical scaling with moderated nonlinear factor analysis. *Journal of Educational Measurement*, 62(4), 929–958. <https://doi.org/10.1111/jedm.70019>
- Tackett, J. L., & Hallquist, M. (2022). The need to grow: Developmental considerations and challenges for modern psychiatric taxonomies. *Journal of Psychopathology and Clinical Science*, 131(6), 660–663. <https://doi.org/10.1037/abn0000751>
- Tiego, J., Martin, E. A., DeYoung, C. G., Hagan, K., Cooper, S. E., Pasion, R., Satchell, L., Shackman, A. J., Bellgrove, M. A., Fornito, A., Abend, R., Goulter, N., Eaton, N. R., Kaczkurkin, A. N., Nusslock, R., & the HiTOP Neurobiological Foundations Work Group. (2023). Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. *Nature Mental Health*, 1(5), 304–315. <https://doi.org/10.1038/s44220-023-00057-5>
- Tyrell, F. A., Yates, T. M., Widaman, K. F., Reynolds, C. A., & Fabricius, W. V. (2019). Data harmonization: Establishing measurement invariance across different assessments of the same construct across adolescence. *Journal of Clinical Child & Adolescent Psychology*, 48(4), 555–567. <https://doi.org/10.1080/15374416.2019.1622124>
- Vaillant, G. E. (2012). *Triumphs of experience: The men of the Harvard Grant Study*. Harvard University Press. <https://doi.org/10.2307/j.ctt2jbxsl>
- Van De Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthen, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, 4(770). <https://doi.org/10.3389/fpsyg.2013.00770>
- Wakschlag, L. S., Briggs-Gowan, M. J., Choi, S. W., Nichols, S. R., Kestler, J., Burns, J. L., Carter,

- A. S., & Henry, D. (2014). Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(1), 82–96.e3. <https://doi.org/10.1016/j.jaac.2013.10.011>
- Wakschlag, L. S., Roberts, M. Y., Flynn, R. M., Smith, J. D., Krogh-Jespersen, S., Kaat, A. J., Gray, L., Walkup, J., Marino, B. S., Norton, E. S., & Davis, M. M. (2019). Future directions for early childhood prevention of mental disorders: A road map to mental health, earlier. *Journal of Clinical Child & Adolescent Psychology*, 48(3), 539–554. <https://doi.org/10.1080/15374416.2018.1561296>
- Walters, G. D. (2011). The latent structure of life-course-persistent antisocial behavior: Is Moffitt's developmental taxonomy a true taxonomy? *Journal of Consulting and Clinical Psychology*, 79(1), 96–105. <https://doi.org/10.1037/a0021519>
- Walters, G. D. (2012). Developmental trajectories of delinquent behavior: One pattern or several? *Criminal Justice and Behavior*, 39(9), 1192–1203. <https://doi.org/10.1177/0093854812443645>
- Walters, G. D. (2014). Continuous versus categorical models of delinquency risk. *American Journal of Criminal Justice*, 39(3), 395–410. <https://doi.org/10.1007/s12103-013-9235-1>
- Walters, G. D. (2015). The latent structure of criminal persistence: A taxometric analysis of offending behavior from late adolescence to early adulthood in adjudicated male delinquents. *American Journal of Criminal Justice*, 40(3), 542–559. <https://doi.org/10.1007/s12103-014-9272-4>
- Walters, G. D., & Ruscio, J. (2013). Trajectories of youthful antisocial behavior: Categories or continua? *Journal of Abnormal Child Psychology*, 41(4), 653–666. <https://doi.org/10.1007/s10802-012-9700-1>
- Walton, K. E., Ormel, J., & Krueger, R. F. (2011). The dimensional nature of externalizing behaviors in adolescence: Evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology*, 39(4), 553–561. <https://doi.org/10.1007/s10802-010-9478-y>

- Ward, C., Oleson, J., Tomblin, J. B., & Walker, E. (2022). Modeling population and subject-specific growth in a latent trait measured by multiple instruments over time using a hierarchical Bayesian framework. *Journal of Applied Statistics*, 49(2), 449–465. <https://doi.org/10.1080/02664763.2020.1817346>
- Weeks, J. P. (2018). An application of multidimensional vertical scaling. *Measurement: Interdisciplinary Research and Perspectives*, 16(3), 139–154. <https://doi.org/10.1080/15366367.2018.1502005>
- Weems, C. F. (2008). Developmental trajectories of childhood anxiety: Identifying continuity and change in anxious emotion. *Developmental Review*, 28(4), 488–502. <https://doi.org/10.1016/j.dr.2008.01.001>
- Weiss, B., & Garber, J. (2003). Developmental differences in the phenomenology of depression. *Development and Psychopathology*, 15(2), 403–430. <https://doi.org/10.1017/S0954579403000221>
- Welzel, C., Brunkert, L., Kruse, S., & Inglehart, R. F. (2023). Non-invariance? An overstated problem with misconceived causes. *Sociological Methods & Research*, 52(3), 1368–1400. <https://doi.org/10.1177/0049124121995521>
- Welzel, C., Kruse, S., & Brunkert, L. (2023). Against the mainstream: On the limitations of non-invariance diagnostics: Response to Fischer et al. And Meuleman et al. *Sociological Methods & Research*, 52(3), 1438–1455. <https://doi.org/10.1177/00491241221091754>
- Whitbourne, S. K. (2019). *Longitudinal, cross-sectional, and sequential designs in lifespan developmental psychology*. Oxford University Press.
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, 4(1), 10–18. <https://doi.org/10.1111/j.1750-8606.2009.00110.x>
- Winter, S. D., & Depaoli, S. (2020). An illustration of Bayesian approximate measurement invariance with longitudinal data and a small sample size. *International Journal of Behavioral Development*, 44(4), 371–382. <https://doi.org/10.1177/0165025419880610>

World Health Organization. (2022). *International Statistical Classification of Diseases and Related Health Problems 11th revision (ICD-11)*. <https://icd.who.int>

Wright, A. G. C., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *Journal of Abnormal Psychology, 122*(1), 281–294. <https://doi.org/10.1037/a0030133>

Yang, Y., & Land, K. C. (2013). *Age-period-cohort analysis: New models, methods, and empirical applications*. Taylor & Francis.

Zhai, R., Zhang, L., & Cheng, G. (2026). Developmental trajectories of emotional and behavioral dysregulation from ages 3 to 17: Early predictors and adolescent developmental outcomes. *Journal of Affective Disorders, 405*, 121556. <https://doi.org/10.1016/j.jad.2026.121556>