Survey Research Methods (2025)
Vol. 19, No. 4, pp. 371-388
https://doi.org/10.18148/srm/2025.v19i3.8442
European Survey Research Association

A matter of perspective? Differences between adolescent-parent and parent-teacher pairs in responses to the Strengths and Difficulties Questionnaire using a Scottish national cohort study

Madison Bunker[1] · Valeria Skafida[1] · Emma Davidson[1]
[1]University of Edinburgh

Although multiple-respondent scoring methods are increasingly recommended for youth mental health questionnaires, utilisation of parent-only responses remains common in survey research. The substantive, epistemological and methodological ramifications of this perspective gap remain under-explored despite the widespread adoption of youth psychometrics in general population social survey datasets. Modelling the impact of respondent-pair identities on inter-respondent discrepancies in youth mental health questionnaires reveals "whose" responses differ and how measurement error may be patterned in single-respondent models. Comparing Goodman's Strengths and Difficulties Questionnaire (SDQ) responses from parents, teachers and adolescents themselves, we apply latent difference score modelling to parent-adolescent (age 14, $n = 2943$) and parent-teacher (age 10, $n = 1833$) pairs from the Growing Up in Scotland birth cohort study and present significant inter-respondent differences in behavioural perceptions between these groups. Higher levels of difficulties are associated with larger inter-respondent discrepancy levels. The impact of gender, housing tenure, finances, family composition, maternal mental health and education on score discrepancies vary in direction, magnitude and significance between SDQ behavioural components. Therefore, discrepancies depend upon characteristics of each measured behaviour, not a global propensity to dis/agreement. Evaluating the implications of these findings, we advocate for the inclusion of youth self-reporting in survey datasets and discuss how to caveat research with the potential impact of respondent identities when multiple respondent data are unavailable.

Keywords: measurement; measurement discrepancies; cohort studies; goodman's strengths and difficulties questionnaire; child and adolescent mental health

1 Introduction

Goodman's (1997) Strength's and Difficulties Questionnaire (SDQ) is a leading measure of children and young people's[1] mental health outcomes in social and economic inequalities research (Reiss 2013). It is prevalent in large-scale datasets; 21 out of 55 British cohort or longitudinal studies include some SDQ coverage (CLOSER 2020), intertwining survey research fields significantly with the measure. Expansive SDQ implementation is reflective of a policy impetus to understand youth mental health through representative population studies (Newlove-Delgado 2023). This is in light of increasing mental health difficulties

Supplementary Information The online version of this article (https://doi.org/10.18148/srm/2025.v19i3.8442) contains supplementary material, which is available to authorized users.

Corresponding author: Madison Bunker, University of Edinburgh, Edinburgh, UK (Email: maddi.bunker@ed.ac.uk)

[1] The SDQ is used from early childhood to late adolescence, and a young adult version is available. For consistency, we refer to "children" as people under 13 years-old, "adolescents" as those 13 to 17, and "youth" or "young people" to encompass both.

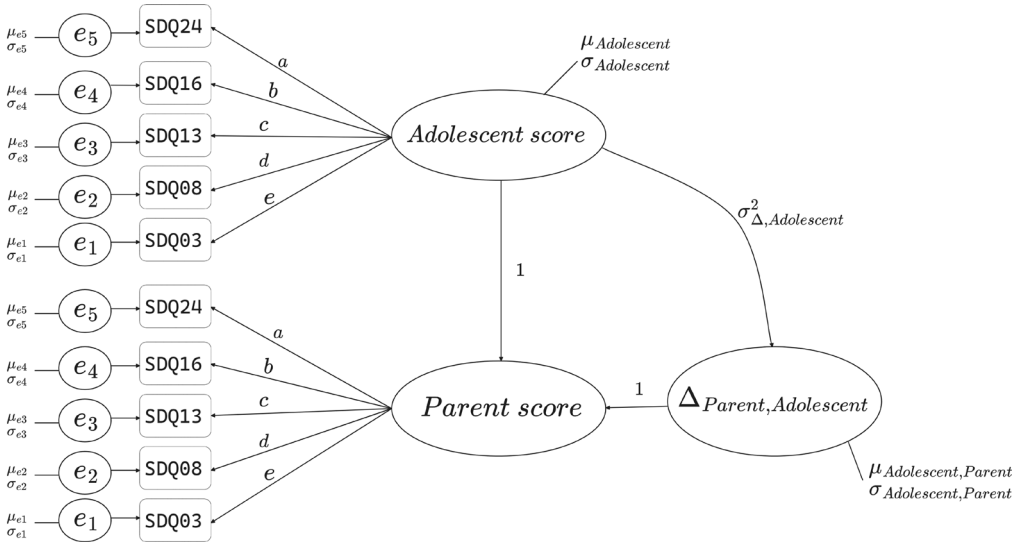

Fig. 1

LDS path diagram of the emotional difficulties component, adapted from (de Haan et al. 2018, pg. 360)

amongst the youngest generation (Armitage et al. 2023; The Lancet 2021). The SDQ's popularity equally necessitates that its characteristics and potential biases in general population surveys are known and confronted in research practice.

Parent, teacher and adolescent self-reporting versions of the SDQ are available. Guidelines both from Goodman et al. (2004) and more recent research Newlove-Delgado (2023) promote combining responses through multi-informant algorithms. These have been found to perform significantly better for identifying psychiatric disorders (Aebi et al. 2017; Goodman et al. 2010; Vugteveen et al. 2021). Furthermore, parents (Kersten et al. 2016) and teachers (White et al. 2013) in qualitative SDQ studies have voiced unease with being the sole authority on a child's behaviour. Participants suggest including other respondents to increase perspective breadth (*ibid*). Nonetheless, many survey datasets—and subsequently, much research—provide greater coverage of parent-reported scores, resulting in an enlarged influence of their perspectives in the literature.

This matters because whose responses are operationalised in youth studies intractably shapes research conclusions (De Los and Kazdin 2005), altering the magnitude and statistical significance of results depending upon the measures and respondents utilised (Johnston et al. 2010). Under-utilisations of young people's voices in particular can be attributed to a form of "epistemic injustice" in research (Carel and Györffy 2014), an often overlooked form of discrimination (Fricker 2007). Comparing discrepancies between the perspectives of adolescents and different adults in young people's lives, and therefore the impact of their specific vantage, is prudent for unpacking these biases. Substantively, inter-respondent discrepancies reveal differences in groups' perceptions of behaviours (De Los et al. 2013). Methodologically, discrepancies in multiple-informant data can shine a light on how respondent characteristics might influence scoring in studies with a single respondent group. As applied survey data researchers, we hope that advancing these two fronts will improve our utilisations of youth data.

Despite waxing interest in inter-respondent differences, consensus has been stymied by methodological concerns with many popular discrepancy models (Laird and De Los 2013). Furthermore, few discrepancy studies have empirically tested for metric invariance (MI), i.e. whether respondent groups respond to the instrument equivalently, in the SDQ. This leaves the source of discrepancies entangled between differences in *perceptions* of behaviours and differences in respondent groups' *interactions* with the questionnaire itself (de Haan et al. 2018; Putnick and Bornstein 2016). These are both important but distinct sources of measurement discrepancies.

Recent findings from the Millennium Cohort Study overcome some limitations using latent difference score (LDS) modelling. Here data evidenced significant discrepancies between parent and adolescent SDQ responses at age 17, and different discrepancy trends depending on parent and adolescent characteristics (Booth et al. 2023). Applying LDS methods to a representative sample of Scottish households, we build upon this work by addressing several re-

search gaps identified in the paper; extending inter-respondent discrepancies to parent-teacher respondent pairs, parent-child pairs in earlier adolescence, and a different national setting using a large scale survey dataset. By combining a sociological vantage with increased coverage of socio-demographic variables, we contribute to discerning the pathways from which measurement discrepancies emerge and discuss the implications of our findings for survey data users.

2 Background

2.1 Perspective Divergence

Beyond "mere measurement error", respondent discrepancies in youth mental health metrics are increasingly understood to constitute substantive differences between informants' judgement or awareness of behaviours (Achenbach 2011). This prompts us to consider how and why parents, teachers and adolescents develop divergent views, and therefore how we may predict the nature of inter-respondent discrepancies in general population samples. Goffman's (1956) theory of the presentation of self is an instructive framework because it demonstrates the role of inter- and intra-personal perspectives within social interaction mechanisms. For Goffman, an individual's attributes are only ever assessed indirectly, first mediated through performance of their daily actions, which are moderated to match the audience and context, and secondarily through the audience's impression of the performance (see Goffman 1956, pg. 10).

How may these interactions and subsequent perceptions systematically differ? Diverse values and life experiences between families of different social, class and ethnic backgrounds influence parent-child relationships and parenting practices (Lansford 2022). Interpersonal interactions are further mediated by individual lenses, such as gender identity (Starrels 1994) or mental state (Lohaus et al. 2020). For example, children may act differently between school and home, or different teachers and parents may interpret the same actions and demeanours discordantly. Both of these processes influence SDQ assessments. Any individual attempt to understand and report a child's strengths and difficulties arises from the respondent's and child's unique combination of these interactions and settings.

Beyond face-to-face interactions, scoring the questionnaire is another social performance to the indirect audience of the researcher. Ergo, reports constitute social interactions which are mediated by the reflexive perceptions of one's backstage, or their "self-concept" (Epstein 1973), and their performance to an imagined audience through the questionnaire. For example, participant-questionnaire interactions may take the form of impression management, where respondents moderate responses to positively reflect themselves or their child (Lilly et al. 2018). Respondent identities are also at play here. One's self concept is formed through social interactions during childhood (Harter 2006), in reference to cultural values and interpersonal relationships (Brummelman and Thomaes 2017), and is structured by one's perception of their social place in reference to others through social structures and inequalities (Gecas 1982; Rosenberg 1990).

All of these identities impact how the behaviours are perceived (the parent or teacher's impression of the performance), and around which settings and audiences behaviours are displayed (the young person's moderation of the performance). Furthermore, divergent performance moderation across settings can be a deliberate tactic individuals employ to navigate their social worlds. For example, people on the autism spectrum may utilise camouflaging or masking strategies to appear more "neurotypical" (Lai et al. 2017). Discrepancies provide a window into how participant characteristics impact these vantages and consequently shape estimates in survey data.

Considering this close relationship between respondent identities, social performances and perceptions thereof, we can hypothesise that the magnitude and direction of inter-respondent differences differs between respondent pairs, resulting in an unquantified measurement uncertainty in single-respondent models that is not equally distributed in the population. Including predictive variables in discrepancy models tests the influence of characteristics which explain or alter respondents' frame of reference, and these impacts on behavioural measurement in survey datasets.

2.2 Predictive Attributes

Implementing this framework, what characteristics predict the manifestation of perspective divergence between parent, adolescent and teacher scoring? Goodman's SDQ was originally developed with schoolchildren in dental and psychiatric clinic waiting rooms in 1990s London and may carry assumptions from this target population (Goodman 1997). There are mixed results for metric performance in international samples (Goodman et al. 2012; Kumar and Fonagy 2013; Palmieri and Smith 2007; Rønning et al. 2004) and groups with special traits and needs (Niclasen and Dammeyer 2016; Stokholm and Lykke 2020). Qualitative research reveals that respondents from different cultural backgrounds view behaviours described in the questionnaire differently, impacting their rating of these questions as a strength or difficulty (Kersten et al. 2016). Furthermore, individual characteristics such as gender identity im-

pact how children and young people view themselves, how they behave in different settings (West and Zimmerman 1987), and how adults such as teachers perceive them (Olsson 2023).

Research has shown that parental attributes such as mental health (Lohaus et al. 2020), educational attainment and country of birth (Boman et al. 2016) influence how they assess their children's behaviours. Furthermore, competing explanations for the relationship between maternal depression and mother-reported child mental health scores persist without consensus on the cause or direction of the relationship (Ordway 2011; Richters 1992). The situation effect model posits that children of mothers with depression may display more problematic behaviours at home than at school, leading to parent-teacher disagreement (Fergusson et al. 1993). Conversely, distortion models attribute differences to mothers' perception biases (Najman et al. 2000; Richters 1992). Previous evaluations of SDQ data support a positive association for less parent-adolescent differences amongst parents with depressive symptoms (Booth et al. 2023), except for the emotional symptoms score (Madsen et al. 2020). Proposed explanations also suggest that teachers may be less reactive to the impact of maternal depression on children and adolescents' difficulties compared to parents (ibid).

Parent-teachers score differences are driven by multiple factors. For example, educators' perceptions of children's behaviours may be influenced by their perceptions of the family attributes, children may act differently in the classroom environment than at home, and the total time spent with the child differs between the two respondent groups (Winterbottom et al. 2008). Family socio-economic status impacts parent-teacher communications (Ankrum 2016) which may foster discordance in how each adult views the child. For example, teachers from the Growing Up in Ireland study rated children attending some centre-based care with higher SDQ-derived total difficulties scores than their peers in parent-only care, but this relationship was not present in parents' scores (Russell et al. 2016). In fact, parents of children attending centre-based care reported fewer peer problems and emotional difficulties compared to their peers (ibid), suggesting that family characteristics impact educators' perceptions in ways which are not reflected in, or may even be oppositional to, these effects on the parents' vantage. Family composition and maternal employment have also been found to be significantly associated with parent-teacher differences across multiple European countries (Cheng et al. 2018). However, these findings do not account for parent-adolescent discrepancies, nor do they explain why these patterns exist or how they apply to different national contexts. Further evidence is needed to disentangle the complex relationship between adolescent and parental attributes, mental health behaviours, and perceptions thereof.

3 Methods

3.1 Data

Data in this paper are drawn from the Growing Up in Scotland (GUS) study. GUS is a birth cohort of children born between 2004/5 in Scotland and their families, including biannual longitudinal follow-up "sweeps" (Scottish Centre for Social Research 2023). Participant households were selected through stratified sampling to create a nationally representative sample when combined with survey weights to attenuate for the non-equal probability of selection and restore representativeness after longitudinal attrition. We merged SDQ responses from parent reports when study children were 10 years old (2014/15, sweep 8), a survey of the study child's teacher (2014/15) contemporaneous to sweep 8 (Scottish Centre for Social Research 2022), and parent and adolescent self-reports at age 14 (2019/20, sweep 10). The majority of parent respondents in these matched sub-samples are the study child's biological mother (sweep 10 = 94%, sweep 8 = 96%). Our findings may consequently differ from other types of parents, although previous research suggests that SDQ agreement between mothers and fathers is sufficiently high (Fält et al. 2018).

The onset of the COVID-19 pandemic and subsequent lockdown conditions impacted GUS data collection for sweep 10; 82% ($n = 2417$) of cases were collected before March 2020 with face-to-face computer-assisted personal interviewing, leaving 18% ($n = 526$) of the sample collected after the UK national lockdown through web and telephone methods (Scottish Centre for Social Research 2022). These observations were excluded to control for the unknown impact of major changes to interview method and context. Their exclusion does not alter the sample representativeness.

3.2 Variables

We now present the chosen variables and their measurement.

Goodman's SDQ is a 25-item instrument designed to measure four dimensions of psycho-pathological "difficulties"—conduct problems, hyperactivity/inattention, emotional issues and peer problems—and one dimension of pro-social "strengths" (Goodman 2001). Each component sums five Likert items about the young person's recent

behaviours, and can be interpreted as discrete factors or combined into a multidimensional additive score of "internalising" and "externalising", or "total difficulties" and prosocial propensities (Goodman and Goodman 2009). This paper includes parent and teacher SDQ responses around age 10, and parent and adolescent responses around age 14.

Respondent and household attributes were also selected to predict perspective differences. Due to non-identical questionnaires at data collection sweeps 8 (2014/15) and 10 (2019/20), and different hypotheses for parental discrepancies between teachers and adolescents, some predictive variables differ between parent-adolescent and parent-teacher pairs.

Gender is measured by self-identification for adolescent-parent pairs at age 14 (sweep 10 = 50% girls). Self-identification is preferential because we are interested in gendered expressions and perceptions thereof, not biological functions (Reisner et al. 2015). Unfortunately, observations were too few to study those who identified their gender as "in another way" ($n = 16$). Additionally, parent-reported child biological sex was used for modelling teacher-parent pairs because gender self-identification was not asked at age 10 (sweep 8 = 52% girls). Although biological sex can be a problematic proxy for gendered experiences (Lindqvist et al. 2021), adults are externally observing the child, affecting their perceptions regardless of whether these assumptions agree with the child's internal experience of gender. We consequently propose that it is reasonable to assume a close relationship between biological sex and adult perceptions of children.

Parental mental health at sweep 10 is measured with a binary response to whether the respondent "ever had period of several days feeling depressed" (sweep 10 = 43% responded affirmatively). Explicit reporting was chosen because it requires individual understandings of one's own health. This question selects parents who recognize their own distress and may assess their child through this lens, thus testing the distortion model. However, the resulting variable is atemporal and identifies almost half of the sample as having experiences of depression, so it may not compare well with other mental health metrics. Equivalent self-assessments are not available for parent-teacher data at sweep 8, but 12-items were collected from the Short Form Mental Health Component (MCS-12) scale, a validated measure of general psychological health (Ware et al. 1998). MCS-12 scores from the main parental respondent were dichotomised at one standard deviation below the sample mean (sweep 8 MCS-12 mean = 39). Those under this this threshold are defined as having poorer-than-average mental health (13% rated below).

Housing tenure is a binary measure of parents who reported to own their residence (sweep 10 = 76% owned residence, sweep 8 = 77%) compared to renting. The family stress model asserts that household stressors, such as housing instability, influence psychological distress and intra-family communication (Masarik and Conger 2017). Tenancy status is also highly associated with socio-economic status and other aspects of disadvantage in Scotland (Rolfe et al. 2020). These interrelated impacts on young people's behaviours and perceptions thereof may diverge between respondent groups, producing non-equivalent discrepancies in the survey population.

Languages spoken at home were used to proxy cultural diversity, a method implemented previously for GUS data (Skafida 2014). Households are dichotomised to those with "only English spoken in the home" (sweep 10 = 95%) or "some language other than English spoken". Families with immigration histories speaking multiple languages may retain cultural differences that impact children's behaviours, adult perceptions of behaviours, and interactions with the cultural assumptions in SDQ questions. A preliminary model used parental country of origin information, but language was chosen to include second generation families and other ties to non-English speaking cultures. This choice did not change the substantive results.

Household finances are the parent's response to "how you and your family are managing financially these days", dichotomised to "managing well" or "other" than managing well (sweep 10 = 61% managing well, sweep 8 = 55%). Subjective measurement was chosen to detect the impact of parental perceptions of their difficulties on scoring.

Family composition asks whether the household has parents living as a couple (sweep 10 = 79% couple families, sweep 8 = 84%) or single parent(s). Single parenthood may influence perceptions of families or alter adult awareness of children's issues because the parenting is not shared in the household.

Parental education indicates if the parent was educated at or above the Scottish Highers level (sweep 8 = 85% Highers or above). We theorised that teachers may perceive children of parents with fewer education qualifications to have greater difficulties due to stigma against these parents.

3.3 Model Choice

This section provides a brief overview and justification of the study methods and some familiarity with statistics is assumed.

Estimating respondent discrepancies with predictive variables poses a methodological conundrum. One must disentangle the relationship between predictive variables and score discrepancies—e.g. how gender impacts differences between parent and adolescent respondents' scores—from the main relationship between predictive variables and item scores—e.g. how gender impacts the scores overall. This

is challenging due to the correlated residual error of these variables (see Edwards 1994; and Laird and De Los Reyes 2013) for a detailed mathematical explanation).

The latent difference score (LDS) method (de Haan et al. 2018) offers a structural equation modelling solution to the limitations of popular discrepancy models. LDS models estimate discrepancies as differences in *latent factors*, the underlying constructs one wishes to study, measured by combining manifest/observed variables. Latent factors are estimated from the shared variance of manifest items, separating them from the residual error of each question (Loehlin and Beaujean 2017). Each SDQ component can be treated as a single latent factor measured by five questions about the behaviour. The LDS model constructs a second-order latent factor equal to the difference between two respondent groups' latent factors, representing the difference in the respondents' perception of the latent behaviour without measurement or residual errors (de Haan et al. 2018).

Figure 1 displays a visual representation of the LDS model for parent-adolescent responses to the SDQ emotional difficulties score, adapted from de Haan et al.'s (de Haan et al. 2018) path diagram. Denoted by repeated letters in the diagram, the factor loadings, intercepts and structure of each latent factor are constrained to be equal between the groups, e.g. parents and adolescents. The result of these constraints is that the LDS score ($\Delta_{Parent,Adolescent}$) is entirely composed of the distance between the two respondent's perceptions of the latent construct.

Building upon Booth et al.'s (Booth et al. 2023) novel implementation of LDS methods to the R language and Goodman's SDQ, we estimated LDS scores for adolescent-parent reporter pairs ($\Delta_{Parent,Adolescent}$) and teacher-parent reporter pairs ($\Delta_{Parent,Teacher}$) using the Lavaan 0.6–16 (Rosseel 2012) package and R version 4.3.1 (R Core Team 2023). Finally, individual LDS models were re-estimated with dichotomous predictive variables to ascertain the impact of socio-demographic characteristics on respondent groups' average estimates, and discrepancies between these.

Following Vieira, Salgueiro, and Smith's (Vieira et al. 2016) comparison of complex survey design corrections in structural equation modelling applications, two estimation methods were implemented: maximum likelihood with robust Huber-White standard errors (MLR) without survey weighting, and bootstrapped ($n = 500$ resamples) pseudo-maximum likelihood (ML) with survey weighting and sampling design corrections using the lavaan.survey (Oberski 2014) package. Survey weights correct for differential selection probabilities for some children, account for non-response bias, and calibrate the achieved sample to match the characteristics of the population. Survey-weighted ML estimates are reported with full results in the Appendix. Equivalent unweighted MLR estimates and code replication files are available as Supplementary Materials.

3.4 Testing Assumptions

First, we empirically tested several model assumptions for both matched subsamples of GUS data. SDQ component—emotional difficulties, peer problems, hyperactivity/inattention issues, conduct problem and pro-social strengths—are theorised to capture a single dimension of latent factor, combining to five dimensions for the total questionnaire (Goodman 1997)[2]. Confirmatory factor analysis (CFA) tests evaluate the questionnaire's factor structure fit for parents, teachers and adolescents in the sample. Model fit was evaluated holistically with scaled comparative fit index (CFI), root mean square error of approximation (RMSEA), standardised root mean squared residual (SRMR). These metrics were chosen over χ^2 due to its bias against larger sample sizes (Putnick and Bornstein 2016; van de Schoot et al. 2012). After achieving insufficient fit in the five-factor model (Appendix Table A1), univariate CFA of the factor structures were conducted for each component, time period and respondent group separately (Table A2). Hyperactivity and conduct component scores achieved poor model fit in both samples (CFI < 0.90, RMSEA < 0.08, SRMR < 0.08), suggesting these questions do not load onto a single, unique latent factor. Consequently, externalising components were excluded from the study for both respondent group pairings.

Comparing latent factors between groups also requires that the scores are equivalent measures of the latent construct, or "Metric Invariant" (MI) (Putnick and Bornstein 2016; Svetina et al. 2020). Constraining factor structure (configural invariance), item loadings (metric invariance) and intercepts (scalar invariance) to be equal between groups ensures that differences in the latent factors are attributable to perceptions of the behaviour itself, not differences in how groups understand and respond to the questions asked, a distinct source of measurement divergence. Residual invariance tests were excluded because residuals do not contribute to the latent factor (Vandenberg and Lance 2000).

Following guidelines prescribed in (van de Schoot et al. 2012), we tested for MI using nested multi-group CFA tests with parents as the reference group for all respondent pairs where the SDQ component had adequate unidimensional factor fit (full MI fit changes reported in Tables A3 and A4).

Standard MI thresholds suggest small allowances of model fit loss between constrained levels, <0.10 change in CFI, <0.015 change in RMSEA or < change 0.10 SRMR (Chen 2007); however; (Rutkowski and Svetina 2014) forward that Δ –0.020 CFI and Δ 0.03 RMSEA estab-

[2] Questionnaire items for each component can be found at https://www.sdqinfo.org/.

Table 1

Paired CFA fit for partial-scalar configurations

Component	Age	CFI	RMSEA	SRMR	*N*
Parent-adolescent pairs					
Pro-social	14	0.964	0.033	0.033	2159
Peer	14	0.947	0.040	0.039	2018
Emotion	14	0.967	0.046	0.034	2211
Parent-teacher pairs					
Pro-social	10	0.936	0.047	0.039	1554
Emotion	10	0.963	0.040	0.035	1545

Parents are set as the base reporter group
Sampling weights applied

lish acceptable MI between configural and metric models for larger numbers of groups in substantial sample sizes (Putnick and Bornstein 2016). recommend that minor divergences in model fit statistics should not prevent otherwise robust findings from publication, but instead be maintained as a research limitation. Taking a holistic approach, we allowed minor divergence from Δ CFI for configural to metric invariance arrangements (Δ –0.017) for parent and adolescent peer problems components due to over all good model fit (CFI 0.0947, RMSEA 0.041, SRMR 0.038) and very close fit between metric and partial scalar model fit (Δ CFI 0.001, Δ RMSEA 0.001, Δ SRMR 0.000). Therefore, we retained the peer problems component for parent-adolescent pairs with the caveat that it demonstrates less robust MI. The peer problem's scale metric configuration fit was significantly worse for parent-teacher pairs and we could not establish MI. This means that we are not confident that the score is measuring the same construct between the two groups. Consequently, we also excluded the parent-teacher model of the peer problems component.

Full scalar invariance was missed for all components with both parent-adolescent and parent-teacher pairs, but partial invariance with acceptable model fit were established through "backwards" freeing of item intercept constraints using modification indices (Putnick and Bornstein 2016; Vandenberg and Lance 2000). Partial invariance is generally considered statistically justified for comparing latent variables when a majority of items remain constrained and theoretical implications of freeing intercepts are considered (Luong and Flake 2022). For parent-adolescent pairs, intercepts were freed for question 1 ("child considers others") of the pro-social component, questions 14 (child is "liked by children") and 23 (child "gets on better with adults") for the peer problems component, and question 16 ("child loses confidence") for emotional symptoms. Freed intercepts for parent-teacher paired models were question 1 ("child considers others") of the pro-social component, and question 8 ("child worries") for the emotional difficulties component (see Table 1).

4 Results

4.1 Latent Difference Scores

Using partial scalar configurations of the SDQ components to construct LDS models, all reporter pairs demonstrated significant inter-respondent discrepancies between ratings of pro-social strengths, peer difficulties and emotional symptoms (see Table 2).

Positive and significant standardised mean LDS estimates indicate that, on average, adolescents self-reported fewer pro-social strengths, and greater emotional symptoms, than their parents perceived. Adolescents also reported fewer peer difficulties relative to their parents. On average, teachers perceived fewer emotional symptoms and fewer pro-social strengths than parents, suggesting tendencies towards the middle of these scales.

Variance between score discrepancies was smallest for parent-teacher pairs rating the emotional symptoms subscale. This suggests that the level of disagreement between pairs of teachers and parents is more consistent, whereas there is greater variability amongst parent-adolescent pairs. Although the average difference is smallest in the parent-teacher pro-social scale, this estimate also has amongst the highest variance between respondent-pairs. Negative covariances across the board indicate that higher scores predict larger discrepancies for both strengths and difficulties. Significant and large variances between respondent pairs further justifies a need to understand the participant characteristics which predict these non-uniform divergences.

4.2 Conditional Models

Socio-demographic attributes indeed influenced scoring divergently across respondent pairs. Figures 2 through Fig. 6 plot the model-implied means for LDS models conditional on participant characteristic, and Appendix Tables A5 to A16 provide the full numeric estimates for both the model-implied means and the main effects of each conditional variable on the latent difference factor (Δ) (see Supplement tables 2 through 32 for fully standardised and unweighted estimates). Gender, parental mental health, housing tenure, household financial stress and single parenthood each impacted the latent difference factor (Δ) for at least one component at the 95% confidence level.

Considering gender, parents and teachers viewed boys to display fewer pro-social strengths, on average, than girls.

Table 2

Latent different scores

		Estimates			Fit measures			
Component	Age	Mean	Variance	Covariance	CFI	RMSEA	SRMR	*N*
Parent-adolescent pairs								
Pro-social	14	0.627*	0.109*	–0.613*	0.969	0.034	0.033	2159
Peer	14	0.182*	0.111*	–0.263*	0.955	0.026	0.041	2018
Emotion	14	–0.572*	0.128*	–0.639*	0.968	0.032	0.035	2211
Parent-teacher pairs								
Pro-social	10	0.181*	0.130*	–0.751*	0.933	0.040	0.039	1554
Emotion	10	0.307*	0.055*	–0.475*	0.940	0.040	0.042	1545

Parents are set as the base reporter group
Sampling weights applied
* $p < 0.05$

Average discrepancies in pro-social behaviour ratings between parent-teacher ($\Delta_{parent,teacher}$) and parent-adolescent ($\Delta_{parent,adolescent}$) respondent pairs were also significantly greater for ratings of boys than girls. Boy-parent respondent pairs also had elevated discrepancies in the peer problems component, although the difference between boys and girls average self-ratings of peer problems was not significant. Both parents and adolescents perceived worse emotional difficulties for girls than boys at age 14. The average emotional difficulties discrepancy factor was also significantly higher for girl-parent respondent pairs than boy-parent. Contrastingly, distances between boys and girls' scores were not significantly divergent for parent or teacher ratings of emotional difficulties at age 10. The child's gender also did not have a statistically significantly impact on parent-teacher discrepancies in perceptions of emotional difficulties.

Adolescent children of parents with a history of depression reported relatively fewer pro-social strengths, but the average inter-respondent discrepancy level was not significantly different from other respondent pairs. However, parental mental health did impact discrepancies in ratings for emotional difficulties and peer problems. Parents with a history of depression viewed their children to have elevated difficulties relative to average peers' ratings and the parent-adolescent peer problems discrepancy level was significantly higher for parents with a history of depression. However, parent-adolescent discrepancies in emotional difficulties rating were significantly *lower* than for parents who did not report a history of these struggles. This is an inverse of the impact of parent-teacher discrepancies, which were elevated for parents with below average mental health.

Children and adolescents living in rented accommodation were perceived by teachers and themselves to have significantly elevated difficulties relative to their peers in owner-occupied homes. However, the impact of housing tenure on the discrepancy factor was only significant for parent-adolescent responses to the peer problems component. Children from language diverse and English-only speaking homes displayed substantial overlap in mean scores for all respondent groups and SDQ components, with no significant differences. Additionally, differences between parental educational attainment and parent-teacher discrepancies in strengths or difficulties were not significant.

Adolescents from single-parent households rated themselves to have significantly fewer strengths and greater difficulties, on average, than their peers. However, the conditional effect on parent-adolescent score discrepancies were not statistically significant. Single parent households also had greater parent-teacher discrepancies for rating emotional difficulties than couple families, but no other effect path was significant.

Adolescents and children in households which reported subjective financial struggles were assessed by all groups to have elevated difficulties and fewer pro-social behaviours relative to their peers, but inter-respondent discrepancies levels were not elevated. In fact, parent-adolescent discrepancies in emotional difficulties ratings were significantly higher for parents who reported to be managing their finances well, compared to those who did not.

The conditional effect of financial difficulties on parent-adolescent discrepancies (Supplement Tables 17–18) and the conditional effect of family composition on parent-teacher discrepancies (Supplement Tables 33–34) were only significant for the weighted ML estimation procedure, but were below the 95% confidence level for MLR estimated figures. Therefore, these results should be treated with more caution.

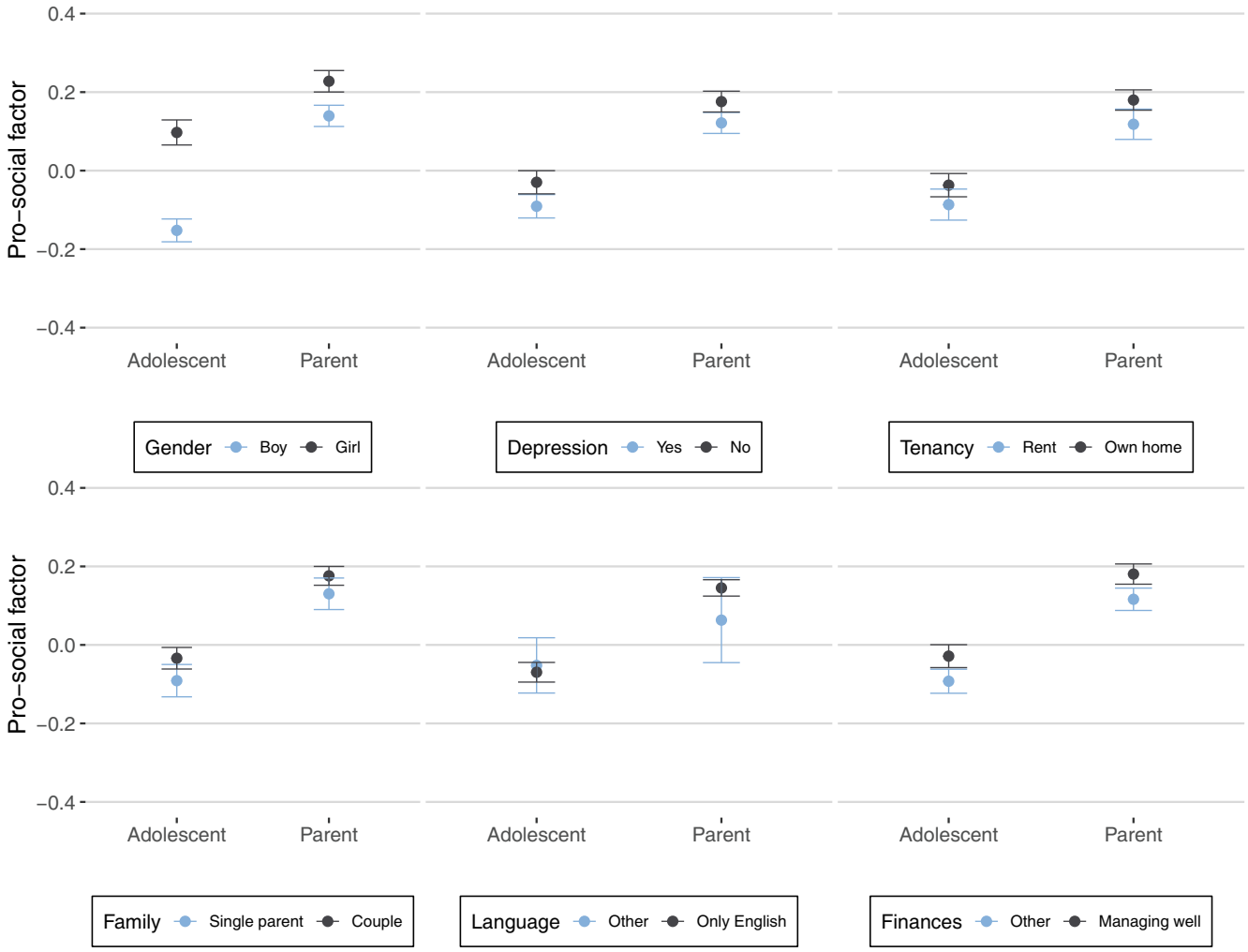

Fig. 2

Model estimated mean latent pro social factor

5 Discussion

Our results demonstrate a substantial influence of respondent attributes on inter-respondent discrepancies between adolescents, parents and teachers' behavioural perceptions across dimensions of social and emotional health. MI constraints ensure that these discrepancies are separated from differing responses to the instrument itself, and should be regarded as average divergence between perceptions of the social and emotional behaviours themselves. Adjacently, MI failure on externalising components reveals a misfit between the theoretical and observed factor structure, highlighting that there is a probable impact of perception *and* measurement divergence in this general population sample which remains unknown.

The association between elevated scores and larger inter-respondent discrepancies suggests that some of the most "extreme" cases are also the most impacted by perspective bias. Consequently, estimates utilising only single-respondent group data are generally missing a broader range of information for those children and adolescents experiencing the greatest difficulties, and the most positive pro-social attributes. Characteristics of inter-respondent divergences are unique in direction and magnitude to each latent construct. This result is generally supported by reports from previous studies which investigate the SDQ at the component, rather than aggregate, level (Booth et al. 2023; Madsen et al. 2020). These patterns imply that discrepancies illustrate divergences on specific dimensions of behaviours, rather than general dis/agreement.

Predictors for bigger score discrepancies do not map neatly onto typical dimensions of disadvantage, particularly for parent-adolescent pairs. This appears to be because some associates of perceiving greater youth difficulties,

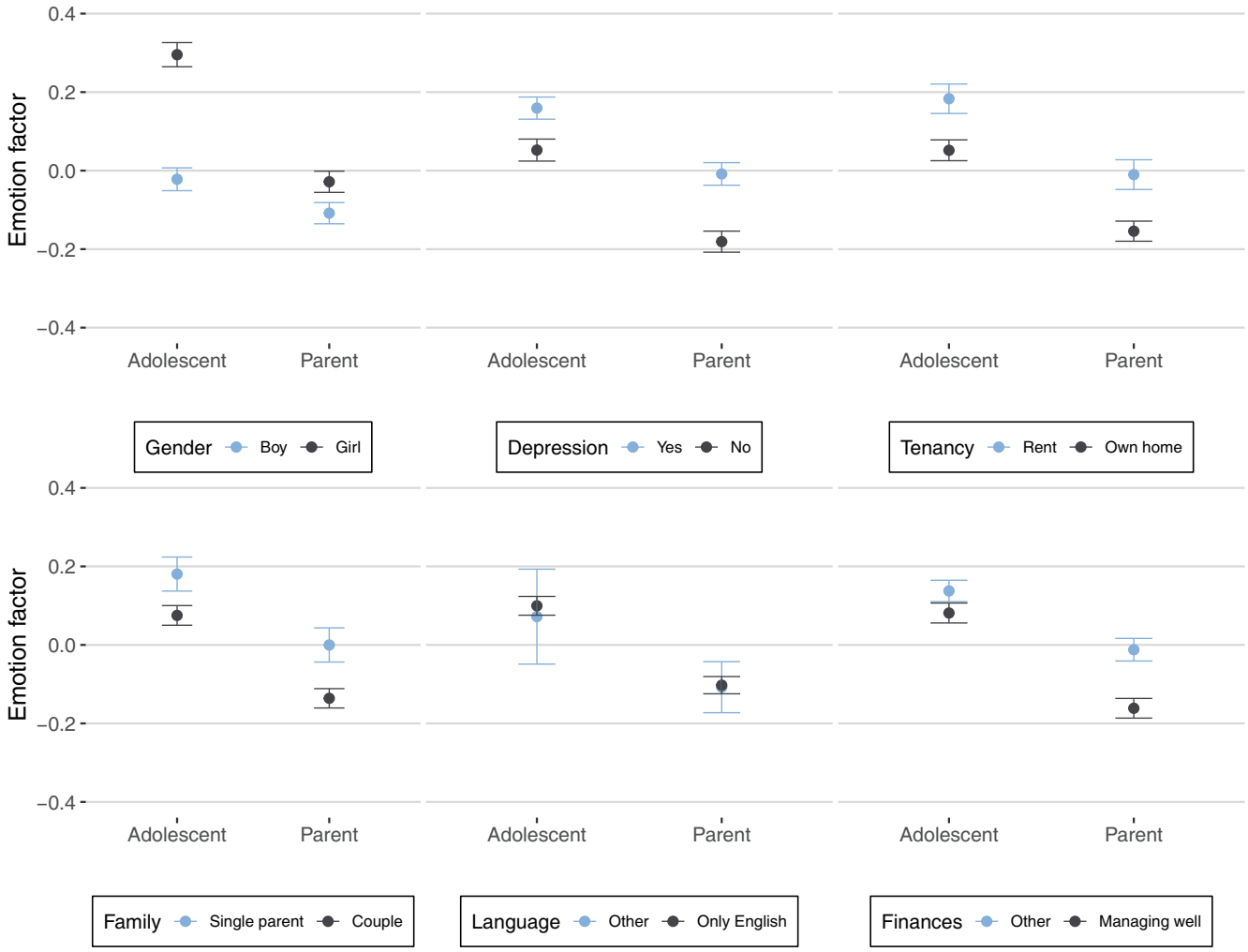

Fig. 3

Model estimated mean latent emotion factor

such as financial struggles, maternal mental ill-health or less-secure housing, seem to have a relatively larger effect on parental perceptions than self-reporting. In some cases this leads to reduced parent-adolescent discrepancies for more disadvantaged groups, such as ratings of emotional difficulties for families reporting financial difficulties.

The nuance of this interplay highlights the value of LDS modelling to disentangle the main effects of predictive characteristics from respondent discrepancy effects. For example, adolescents of parents with a history of depression have larger discrepancies to their parents' scoring in their peer problems estimates, but a smaller difference in their emotional difficulties scores relative to the reference group. The confluence of these competing effects on scoring matters may obfuscate patterns when multiple dimensions of youth mental health are aggregated, such as common usage of SDQ "total difficulties" ratings in survey research applications. The oppositional direction of these results may offer insight into the lack of consensus on the role of parental depression on their children's psychometric scoring (Ordway 2011), because effects are conditional to the dimension(s) of mental health measured.

Contrary to reports in later adolescence (Booth et al. 2023), parents rated significantly greater worries about peer issues than adolescents self-reported. This inter-respondent discrepancy is larger for boys, parents with a history of depression, those living in rented housing, and single-parent households. Parents witness fewer peer-to-peer behaviours in adolescence, so differences may arise from global views on their child from their social vantage, rather than directly observing their performance of social roles. For example, parents often fear that their children conceal bullying outside of the home (Stives et al. 2021), which may heighten their perceptions of peer issues. Additionally, results could

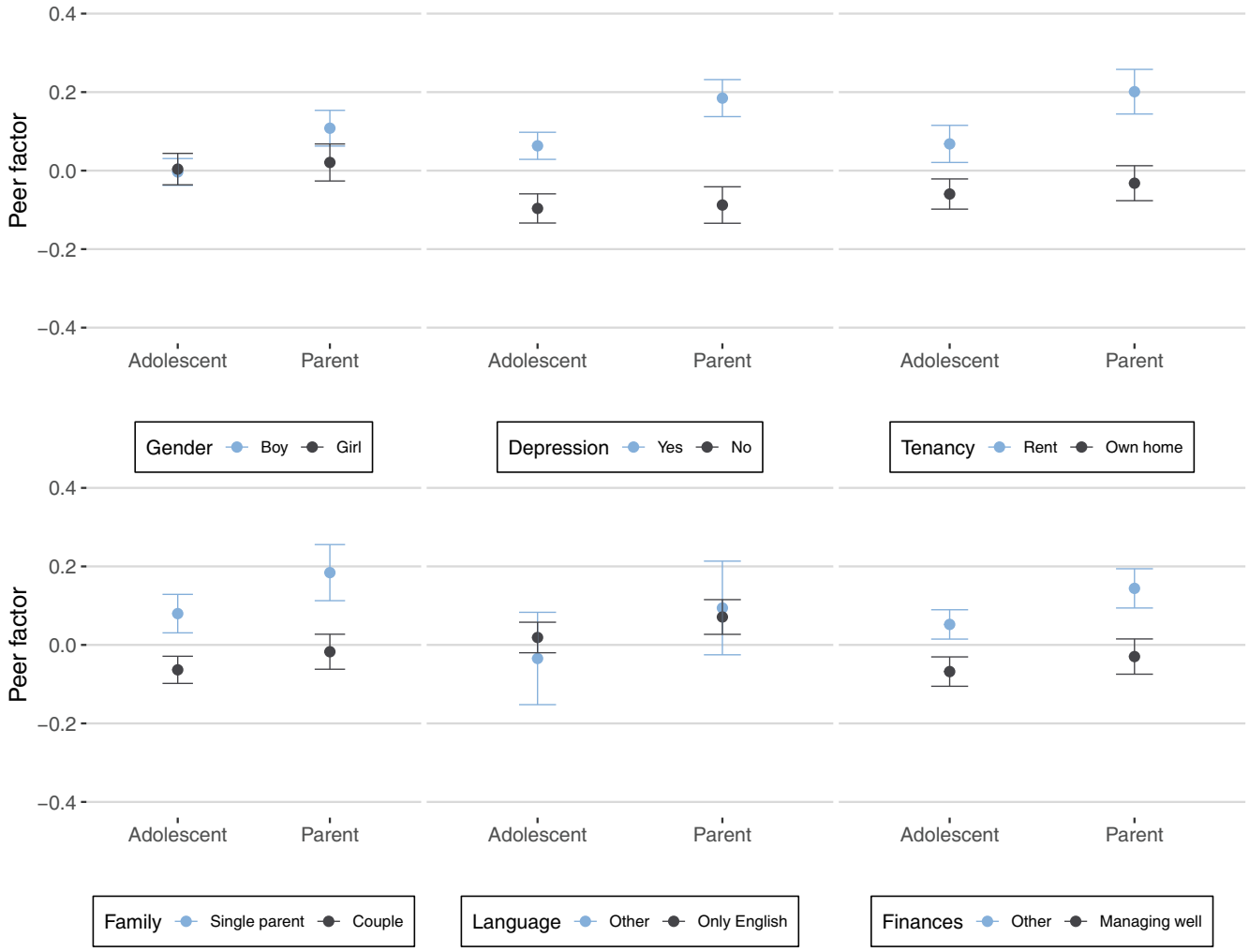

Fig. 4

Model estimated mean latent peer factor

suggest that the parental lens is more influenced by concerns over home instability or disadvantage than self-ratings, or that adolescents are more shielded from difficult circumstances. Differential results from previous research could indicate that this effect is heightened in earlier adolescence, and requires further investigation across other population samples.

Gender had the clearest conditional effect on perspective divergence, predicting large parent-adolescent and parent-teacher discrepancies in the emotional symptoms of girls and the pro-social behaviours and peer problems of boys. Worse emotional symptoms for girls are persistent in general population data (Yoon et al. 2023), and may be understood through the manifestations of biological, psychological and social pressures on the gendered conditioning of child behaviours, leading to heightened expression of emotional symptomology (Chaplin 2015). Discrepancies suggest that social and emotional effects remain under-estimated in parent-reported data compared to multiple respondent perspectives. Furthermore, heightened parental concerns for the peer problems of adolescent boys relative to equal self-ratings may reflect increasing adult concerns about boys' social skills and relationships (Bragg and Ringrose 2022).

Educators' tendencies away from higher scores appear consistent across survey samples at the observed, not latent factor, level (Russell et al. 2016; Winterbottom et al. 2008). Theorising the impact of social performances as a discrepancy pathway, elevated self-reported emotional difficulties may represent a masking of the "performance of self" if adolescents hide their difficulties from parents and teachers through impression management (Goffman 1956; Schlenker and Pontari 2000) or camouflaging in classroom settings (Halsall et al. 2021). Likewise, the contextual effect of

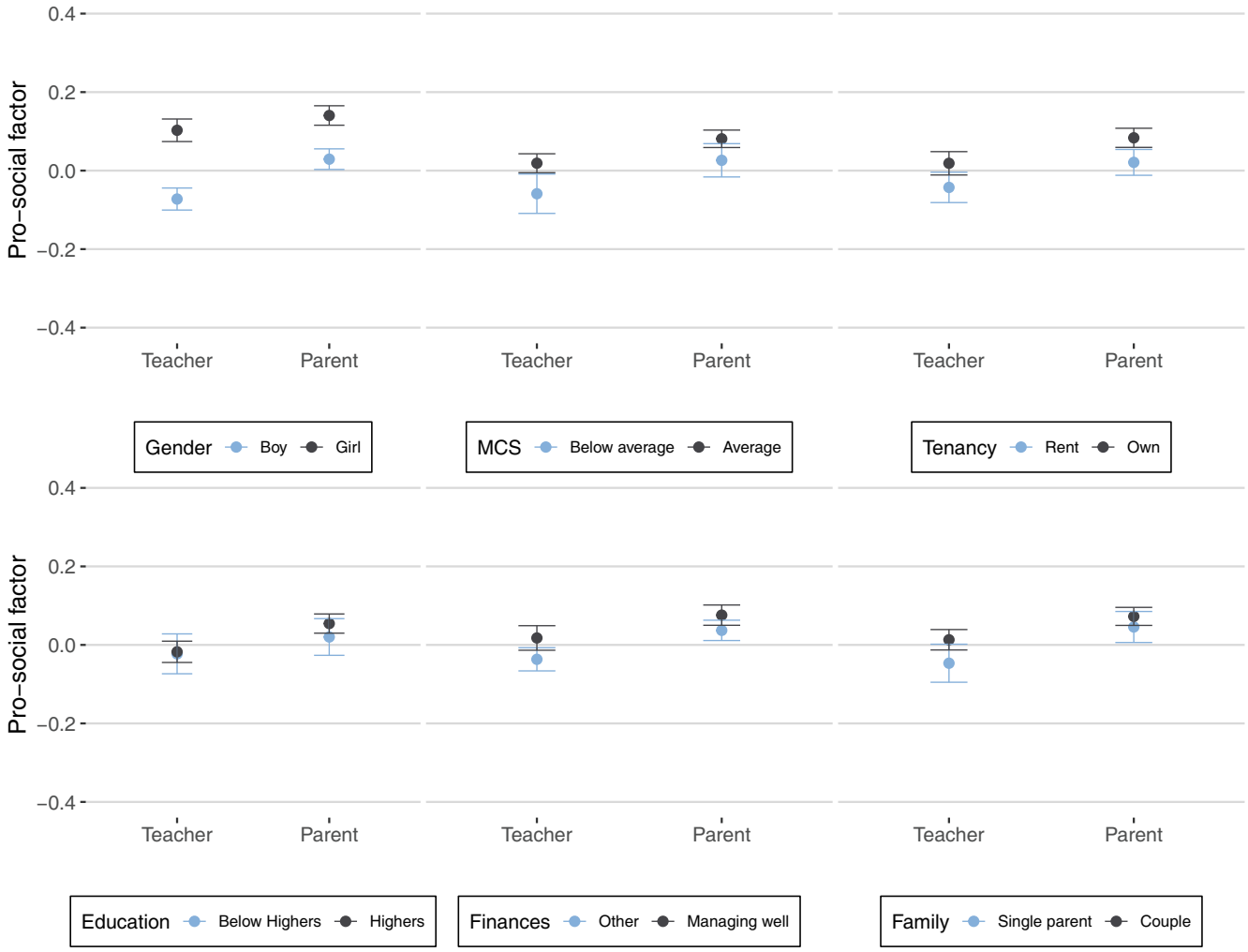

Fig. 5

Model estimated mean latent pro social factor

a classroom setting, reduced time spent with the child, and a greater number of peers the child is compared to, may be tempering teachers' reports. These results indicate that research that relies upon teacher responses alone are likely to be under-estimating strengths and difficulties, relative to how parents or adolescents may be expected to respond.

Although parent-adolescent and parent-teacher models arise from different age groups and are thus not directly comparable in our study, we hypothesise that larger discrepancies among parent-adolescent than parent-teacher pairs may be indicative of an adult-centric bias of the SDQ questionnaire. SDQ questions are external assessment of behaviours (e.g. "*child fidgets or squirms*" or "*child is liked*") which are translated to the adolescents' perceptions of their external behaviours for the self-assessment version (e.g. "*I am restless*" or "*other people generally like me*"). This orientation requires the adolescent to tap into their self-perceptions (Gecas 1982) to understand external perceptions of their actions across multiple settings. Impacts of this self-reflexive practice may contribute to the the poor factor fit of externalising component dimensions, and warrants closer investigation. Children and young people's subjective experiences of taking the questionnaire, and how they view the behaviours examined, may fill in these qualitative gaps to help explain the empirical results presented here.

6 Conclusion

6.1 Recommendations

The respondent identity effects explored in this paper evidence that divergent perceptions of children and young

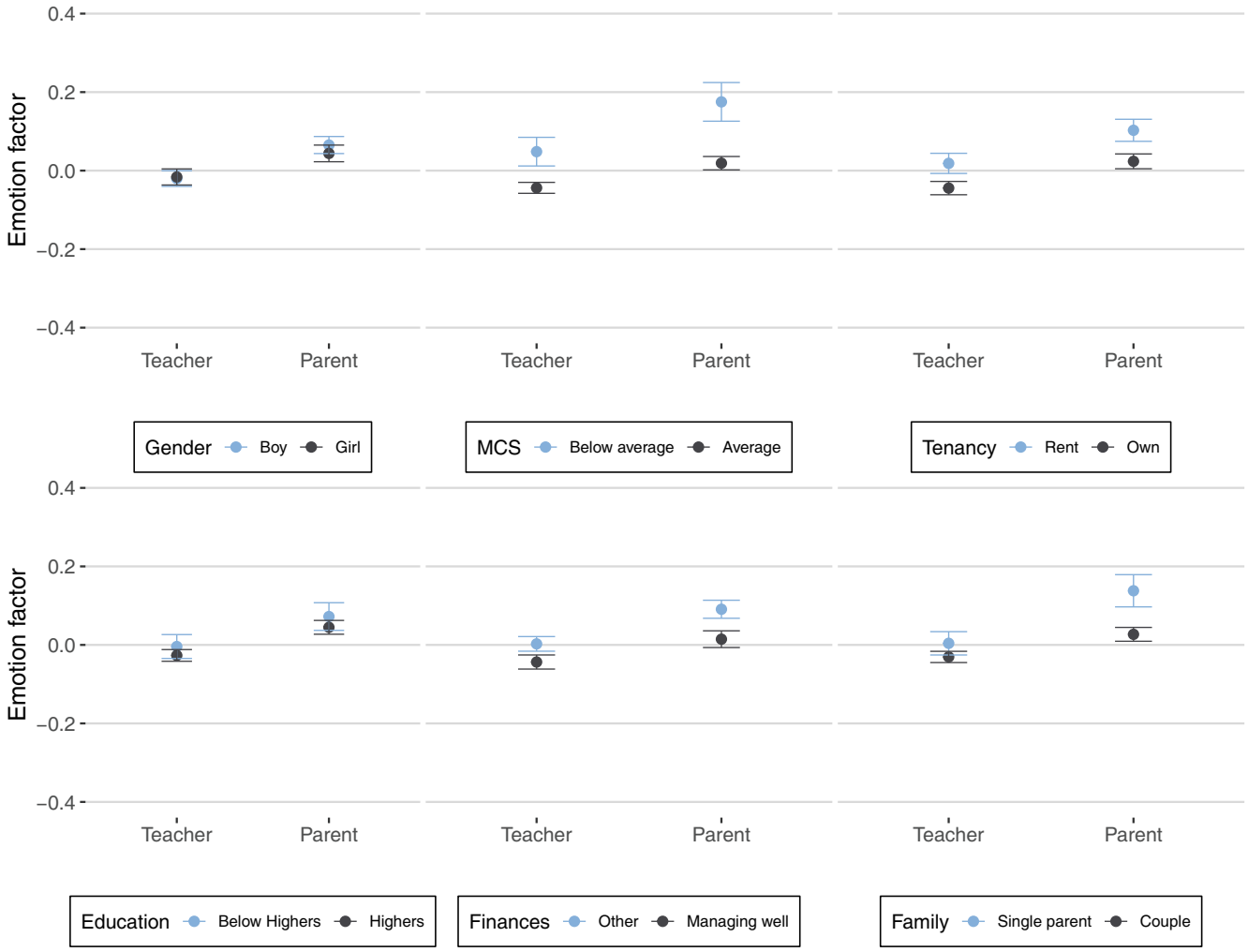

Fig. 6

Model estimated mean latent emotion factor

people's behaviours in survey responses are stratified by participant characteristics, producing a patterned distribution of inter-respondent discrepancies in mental health estimates from survey datasets. This indicates not only that the respondent group chosen will impact estimates of young people's social and emotional behaviours, but that the nature of this impact is non-equivalent across socio-demographic characteristics. Discrepancies are greater for higher behavioural scores, indicating that respondent identities may have a heightened impact on estimating the children and young people that are often of greatest interest to researchers.

Divergence does not require conceptualising any one group's perspective as "incorrect"; instead, participant responses constitute a distribution of relevant but non-interchangeable knowledge on the questions posed. Respondent group differences are a particularly important axis of measurement variability to grapple with in youth research due to the common usage of external informants. Diversity of respondents is a practical solution for addressing this disagreement head-on without favouring any single perspective. We advocate expanding informant variability (Ederer 2004) through multi-respondent scoring to capture a greater range of perspectives (see Bauer et al. 2013; Fält et al. 2017), with particular impetus on closing the representation gap for self-reporting from children and young people.

Where multiple respondent data are unavailable, we suggest that results should be attenuated or interpreted with an understanding of the potential impact of respondent identity on the probable distribution of SDQ scores through critical interpretation and measured evaluation of difference score outcomes. This orientation asks the researcher to put the missing perspective in conversation with the model results. Taking seriously the epistemic injustice of the missing

youth voice in research, we forward that including adolescent perspectives is essential to this exercise. Considering the patterns presented here, what might this missing perspective say?

In general, studies using only teacher data should note that this may underestimate both strengths and difficulties, and studies using only parent data should include the caveat that their results may under-report emotional difficulties and over-report pro-social strengths, relative to adolescent self-reports.

Prior knowledge is also expanded by considering the attributes associated with elevated inter-respondent difference scores. It would therefore be appropriate to apply increased uncertainty around estimates according to attributes associated with greater inter-respondent discrepancies. Because conditional effects vary in magnitude and direction depending upon the component in question, we suggest that attenuation of single-respondent group scores should be implemented at the SDQ component, rather than aggregate total difficulties, level.

Applied survey researchers can weave missing-perspective limitations into their discussion of findings from single respondent-group SDQ data. For example, when using parent-only SDQ responses, researchers should note that parents tend to overestimate girls' emotional difficulties compared to girls own assessments of these. Conversely, parents overestimate boy's pro-social behaviours relative to boys themselves. Given that we find larger discrepancies between parents and children, specifically for emotional difficulties among girls, and for pro-social and peer estimates for boys, researchers should note that parent-only SDQ responses in these domains may be less attuned to children's own perceptions of their behaviour across these dimensions. Similarly, where teacher SDQ reports are being used, researchers should note that relative to parents, teachers under-report emotional difficulties in children in cases where parents have poor mental health, and in single parent households.

We believe that bringing these results into the evaluation of model results from single-respondent group SDQ results allow for more robust and nuanced usage of Goodman's SDQ in general population survey data.

6.2 Limitations and Extensions

Further methodological work is needed to evaluate discrepancies in the externalising component scores, which were excluded from this study due to poor unidimensional CFA fit and failure to establish inter-respondent MI (Booth et al. 2023) were also unable to analyse the hyperactivity/inattention component in their study of adolescent and parent pairs from the Millennium Cohort Study, suggesting the component's multidimensionality may be shared across age, respondent group and regional sample differences in the context of UK general population surveys. We suggest empirically investigating these sources of multidimensionality and metric invariance in the SDQ. Promising methods for this avenue may include reconfiguring non-invariant components through alignment procedures (Asparouhov and Muthen 2014), and qualitative evaluations of participant experiences.

MI failure for some components also warrants normative inspection. For example, lack of scalar invariance between parents and adolescents for item 23 of the "peer problems" component, "child gets on better with adults" may be influenced by discrepancies in viewing this behaviour as a difficulty. New Zealander parents raised concerns with this question on the grounds that, in their culture, this display of family closeness would be a positive behaviour (Kersten et al. 2016). Understanding why respondent groups diverge in responses to the SDQ instrument, just as we investigate divergence in their behavioural perceptions in this work, is integral to incorporating more holistic estimates of mental health behaviours from the questionnaires and responses already available across popular survey datasets. We hope this that is work is useful to researchers for better utilising the data we have (often single-perspectives, particularly in the early years) and informing best practices for advocating the expansion of multiple-respondent information as we design social surveys for the future.

Acknowledgements We would like to thank Professor Lindsay Patterson for providing comments on an earlier version of this paper and the United Kingdom Economic and Social Research Council (ESRC) for funding this research. Grant reference number: 2571396.

References

Achenbach, T.M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child and Adolescent Psychology*, *40*(1), 80–86. https://doi.org/10.1080/15374416.2011.533416.

Aebi, M., Kuhn, C., Banaschewski, T., Grimmer, Y., Poustka, L., Steinhausen, H.-C., & Goodman, R. (2017). The contribution of parent and youth information to identify mental health disorders or problems in adolescents. *Child and Adolescent Psychiatry and Mental Health*, *11*(1), 23. https://doi.org/10.1186/s13034-017-0160-9.

Ankrum, R. (2016). Socioeconomic status and its effect on teacher/parental communication in schools. *Journal of Education and Learning*, *5*(1), 167. https://doi.org/10.5539/jel.v5n1p167.

Armitage, J.M., Kwong, A.S.F., Tseliou, F., Sellers, R., Blakey, R., Anthony, R., & Collishaw, S. (2023). Cross-cohort change in parent-reported emotional problem trajectories across childhood and adolescence in the UK. *The Lancet Psychiatry*. https://doi.org/10.1016/S2215–0366(23)00175-X.

Asparouhov, T., & Muthen, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495–508. https://doi.org/10.1080/10705511.2014.919210.

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*, 475–493. https://doi.org/10.1037/a0032475.

Boman, F., Stafström, M., Lundin, N., Moghadassi, M., Törnhage, C.-J., & Östergren, P.-O. (2016). Comparing parent and teacher assessments of mental health in elementary school children. *Scandinavian Journal of Public Health*, *44*(2), 168–176. https://doi.org/10.1177/1403494815610929.

Booth, C., Moreno-Agostino, D., & Fitzsimons, E. (2023). Parent-adolescent informant discrepancy on the strengths and difficulties questionnaire in the UK millennium cohort study. *Child and Adolescent Psychiatry and Mental Health*, *17*(1), 57. https://doi.org/10.1186/s13034-023-00605-y.

Bragg, S., & Ringrose, J. (2022). *The state of UK boys understanding and transforming gender in the lives of UK boys*. Washington, DC: Global Boyhood Initiative.

Brummelman, E., & Thomaes, S. (2017). How children construct views of themselves: a social-developmental perspective. *Child Development*, *88*(6), 1763–1773.

Carel, H., & Györffy, G. (2014). Seen but not heard: Children and epistemic injustice. *The Lancet*, *384*(9950), 12561257. https://doi.org/10.1016/S0140-6736(14)61759-1.

Chaplin, T.M. (2015). Gender and emotion expression: a developmental contextual perspective. *Emotion review :journal of the International Society for Research on Emotion*, *7*(1), 14–21. https://doi.org/10.1177/1754073914544408.

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464504. https://doi.org/10.1080/10705510701301834.

Cheng, S., Keyes, K.M., Bitfoi, A., Carta, M.G., Kog, C., Goelitz, D., & Kovess-Masfety, V. (2018). Understanding parent-teacher agreement of the Strengths and Difficulties Questionnaire (SDQ): Comparison across seven European countries. *International Journal of Methods in Psychiatric Research*, *27*(1), e1589. https://doi.org/10.1002/mpr.1589.

CLOSER (2020). The Catalogue of Mental Health Measures: Discovering the depths of mental health data in UK longitudinal studies. https://www.cataloguementalhealth.ac.uk/

De Los, R.A., & Kazdin, A.E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*, 483–509. https://doi.org/10.1037/0033-2909.131.4.483.

De Los, R.A., Thomas, S.A., Goodman, K.L., & Kundey, S.M. (2013). Principles underlying the use of multiple informants' reports. *Annual review of clinical psychology*, *9*, 123–149. https://doi.org/10.1146/annurev-clinpsy-050212-185617.

Ederer, E.M. (2004). Mental health problems in young children: Self-reports and significant others as informants. *Psychology Science*, *46*(Suppl 1), 123–140.

Edwards, J.R. (1994). The study of congruence in organizational behavior research: critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*(1), 51–100. https://doi.org/10.1006/obhd.1994.1029.

Epstein, S. (1973). The self-concept revisited: or a theory of a theory. *American Psychologist*, *28*(5), 404–416. https://doi.org/10.1037/h0034679.

Fält, E., Wallby, T., Sarkadi, A., Salari, R., & Fabian, H. (2017). Inter-rater agreement between parent and teacher SDQ ratings in Swedish 3-5-year-olds. *European Journal of Public Health*, *27*(suppl_3), ckx187.136. https://doi.org/10.1093/eurpub/ckx187.136.

Fält, E., Wallby, T., Sarkadi, A., Salari, R., & Fabian, H. (2018). Agreement between mothers', fathers', and teachers' ratings of behavioural and emotional problems in 3-5-year-old children. *PloS One*, *73*(11), e206752. https://doi.org/10.1371/journal.pone.0206752.

Fergusson, D.M., Lynskey, M.T., & Horwood, L.J. (1993). The effect of maternal depression on maternal ratings of child behavior. *Journal of Abnormal Child Psychology*, *27*(3), 245–269. https://doi.org/10.1007/BF00917534.

Fricker, M. (2007). *Epistemic injustice: power and the ethics of knowing*. https://doi.org/10.1093/acprof:oso/9780198237907.001.0001.

Gecas, V. (1982). The self-concept. *Annual Review of Sociology*, *8*, 1–33.

Goffman, E. (1956). *The presentation of self in everyday life*. New York: Knopf Doubleday Publishing.

Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*(11), 1337–1345. https://doi.org/10.1097/00004583-200111000-00015.

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry*, *48*(4), 400–403. https://doi.org/10.1097/CHI.0b013e3181985068.

Goodman, A., Lamping, D.L., & Ploubidis, G.B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, *38*(8), 1179–1191. https://doi.org/10.1007/s10802-010-9434-x.

Goodman, A., Heiervang, E., Fleitlich-Bilyk, B., Alyahri, A., Patel, V., Mullick, M.S.I., & Goodman, R. (2012). Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. *Social Psychiatry and Psychiatric Epidemiology*, *47*(8), 1321–1331. https://doi.org/10.1007/s00127-011-0440-2.

Goodman, R., Ford, T., Corbin, T., & Meltzer, H. (2004). Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *European Child & Adolescent Psychiatry*, *73*(Suppl 2), II25–31. https://doi.org/10.1007/s00787-004-2005-3.

de Haan, A., Prinzie, P., Sentse, M., & Jongerling, J. (2018). Latent difference score modeling: a flexible approach for studying informant discrepancies. *Psychological Assessment*, *30*(3), 358–369. https://doi.org/10.1037/pas0000480.

Halsall, J., Clarke, C., & Crane, L. (2021). "Camouflaging" by adolescent autistic girls who attend both mainstream and specialist resource classes: Perspectives of girls, their mothers and their educators. *Autism*, *25*(7), 2074–2086. https://doi.org/10.1177/13623613211012819.

Harter, S. (2006). The Self. In *Social, Emotional and Personality Development* 6th edn. Handbook of Child Psychology, (Vol. 3, pp. 505–570). Hoboken: Wiley.

Johnston, D., Propper, C., Pudney, S., & Shields, M. (2010). Is there an income gradient in child health? It depends whom you ask. *SSRNElectronic Journal*. https://doi.org/10.2139/ssrn.1575883.

Kersten, P., Dudley, M., Nayar, S., Elder, H., Robertson, H., Tauroa, R., & McPherson, K.M. (2016). Crosscultural acceptability and utility of the strengths and difficulties questionnaire: Views of families. *BMC Psychiatry*, *76*(1), 347. https://doi.org/10.1186/s12888-016-1063-7.

Kumar, M., & Fonagy, P. (2013). The cross-cultural sensitivity of the strengths and difficulties questionnaire (SDQ): a comparative analysis of Gujarati and British children. *International Psychiatry*, *70*(2), 42–45. https://doi.org/10.1192/S1749367600003763.

Lai, M.-C., Lombardo, M.V., Ruigrok, A.N., Chakrabarti, B., Auyeung, B., Szatmari, P., & Baron-Cohen, S. (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism*, *27*(6), 690–702. https://doi.org/10.1177/1362361316671012.

Laird, R.D., & De Los, R.A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: why difference scores cannot Tell you what you want to know and how polynomial regression May. *Journal of Abnormal Child Psychology*, *47*(1), 1–14. https://doi.org/10.1007/s10802-012-9659-y.

Lansford, J.E. (2022). Cross-cultural similarities and differences in parenting. *Journal of child psychology and psychiatry, and allied disciplines*, *63*(4), 466–479. https://doi.org/10.1111/jcpp.13539.

Lilly, M., Davis, T.E., Castagna, P.J., Marker, A., & Davis, A.B. (2018). Maternal impression management in the assessment of childhood depressive symptomatology. *Behavioural and Cognitive Psychotherapy*, *46*(5), 554–569. https://doi.org/10.1017/S1352465818000097.

Lindqvist, A., Senden, M.G., & Renström, E.A. (2021). What is gender, anyway: a review of the options for operationalising gender. *Psychology & Sexuality*, *72*(4), 332–344. https://doi.org/10.1080/19419899.2020.1729844.

Loehlin, J., & Beaujean, A. (2017). *Latent variable models: an Introducation to factor, path and structural equation analysis (fifth)*

Lohaus, A., Rueth, J.-E., & Vierhaus, M. (2020). Cross-informant discrepancies and their association with maternal depression, maternal parenting stress, and mother-child relationship. *Journal of Child and Family Studies*, *29*(3), 867–879. https://doi.org/10.1007/s10826-019-01625-z.

Luong, R., & Flake, J.K. (2022). Measurement invariance testing using confirmatory factor analysis and alignment optimization: a tutorial for transparent analy-

sis planning and reporting. *Psychological Methods*. https://doi.org/10.1037/met0000441.

Madsen, K.B., Rask, C.U., Olsen, J., Niclasen, J., & Obel, C. (2020). Depression-related distortions in maternal reports of child behaviour problems. *European Child & Adolescent Psychiatry*, *29*(3), 275–285. https://doi.org/10.1007/s00787-019-01351-3.

Masarik, A.S., & Conger, R.D. (2017). Stress and child development: a review of the family stress model. *Current Opinion in Psychology. Relationships and Stress*, *13*, 85–90. https://doi.org/10.1016/j.copsyc.2016.05.008.

Najman, J.M., Williams, G.M., Nikles, J., Spence, S., Bor, W., O'Callaghan, M., & Andersen, M.J. (2000). Mothers' mental illness and child behavior problems: cause-effect association or observation bias? *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*(5), 592–602. https://doi.org/10.1097/00004583-200005000-00013.

Newlove-Delgado, D. T. (2023). Using the Mental Health of Children and Young People in England survey series to understand trends in child mental health: Using the Mental Health of Children and Young People in England survey series to understand trends in child mental health: Challenges and opportunities. In UK Data Service Health Studies User Conference, University College London.

Niclasen, J., & Dammeyer, J. (2016). Psychometric properties of the strengths and difficulties questionnaire and mental health problems among children with hearing loss. *The Journal of Deaf Studies and Deaf Education*, *21*(2), 129–140. https://doi.org/10.1093/deafed/env067.

Oberski, D. (2014). lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, *57*(1), 1–27.

Olsson, Ä. (2023). Teachers' gendered perceptions of attention deficit hyperactivity disorder—a literature review. *European Journal of Special Needs Education*, *38*(3), 303–316. https://doi.org/10.1080/08856257.2022.2076476.

Ordway, M.R. (2011). Depressed mothers as informants on child behavior: methodological issues. *Research in nursing & health*, *34*(6), 520–532. https://doi.org/10.1002/nur.20463.

Palmieri, P.A., & Smith, G.C. (2007). Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a U.S. sample of custodial grandmothers. *Psychological Assessment*, *19*, 189198. https://doi.org/10.1037/1040-3590.19.2.189.

Putnick, D.L., & Bornstein, M.H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developm Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.

R Core Team (2023). *R: A language and environment for statistical computing. Manual*. Vienna: R Core Team.

Reisner, S.L., Conron, K.J., Scout, Baker, K., Herman, J.L., Lombardi, E., & Matthews, A.K. (2015). "Counting" Transgender and gender-nonconforming adults in health research: recommendations from the gender identity in US surveillance group. *TSQ: Transgender Studies Quarterly*, *2*(1), 34–57. https://doi.org/10.1215/23289252-2848877.

Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review. *Social Science & Medicine*, *90*, 24–31. https://doi.org/10.1016/j.socscimed.2013.04.026.

Richters, J.E. (1992). Depressed mothers as informants about their children: a critical review of the evidence for distortion. *Psychological Bulletin*, *112*(3), 485499. https://doi.org/10.1037/0033-2909.112.3.485.

Rolfe, S., Garnham, L., Godwin, J., Anderson, I., Seaman, P., & Donaldson, C. (2020). Housing as a social determinant of health and wellbeing: Developing an empirically-informed realist theoretical framework. *BMC Public Health*, *20*(1), 1138. https://doi.org/10.1186/s12889-020-09224-0.

Rønning, J.A., Mrch, W.-T., Handegaard, B.H., & Sourander, A. (2004). The strengths and difficulties self-report questionnaire as a screening instrument in Norwegian community samples. *European Child & Adolescent Psychiatry*, *13*(2), 73–82. https://doi.org/10.1007/s00787-004-0356-4.

Rosenberg, M. (1990). The self-concept: social product and social force. In *Social psychology*. Routledge.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02.

Russell, H., Kenny, O., & Mcginnity, F. (2016). *Childcare, early education and socio-emotional outcomes at age 5: evidence from the growing up in Ireland study*

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement Invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*(1), 31–57. https://doi.org/10.1177/0013164413498257.

Schlenker, B.R., & Pontari, B.A. (2000). The strategic control of information: impression management and selfpresentation in daily life. In *Psychological per-*

spectives on self and identity (pp. 199–232). https://doi.org/10.1037/10357-008.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740.

Scottish Centre for Social Research (2022). Growing up in Scotland: cohort 1, primary 6 teacher survey, 20142015: special licence access. https://doi.org/10.5255/UKDA-SN-8366-1

Scottish Centre for Social Research (2023). Growing Up in Scotland: Cohort 1, Sweeps 1–10, 2005–2020: Special Licence Access. https://doi.org/10.5255/UKDA-SN-5760-13

Skafida, V. (2014). Change in breastfeeding patterns in Scotland between 2004 and 2011 and the role of health policy. *European Journal of Public Health*, *24*(6), 10331041. https://doi.org/10.1093/eurpub/cku029.

Starrels, M.E. (1994). Gender differences in parent-child relations. *Journal of Family Issues*, *75*(1), 148–165. https://doi.org/10.1177/019251394015001007.

Stives, K.L., May, D.C., Mack, M., & Bethel, C.L. (2021). Understanding responses to bullying from the parent perspective. *Frontiers in Education*, *6*.

Stokholm, J.R., & Lykke, K. (2020). The Strengths and Difficulties Questionnaire is a usable way to address mental health at well-child visits in general practice—a qualitative study of feasibility. *BMC Family Practice*, *27*(1), 126. https://doi.org/10.1186/s12875-020-01156-3.

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group Invariance with categorical outcomes using updated guidelines: an illustration using Mplus and the lavaan/semtools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776.

The Lancet (2021). Adolescent wellbeing in the UK. *The Lancet Child & Adolescent Health*, *5*(10), 681. https://doi.org/10.1016/S2352-4642(21)00284-4.

Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–69. https://doi.org/10.1177/109442810031002.

Vieira, M.D.T., Salgueiro, M.D.F., & Smith, P.W.F. (2016). Investigating impacts of complex sampling on latent growth curve modelling. *Journal of Applied Statistics*, *43*(7), 1310–1321. https://doi.org/10.1080/02664763.2015.1100590.

Vugteveen, J., de Bildt, A., Hartman, C.A., Reijneveld, S.A., & Timmerman, M.E. (2021). The combined self- and parent-rated SDQ score profile predicts care use and psychiatric diagnoses. *European Child & Adolescent Psychiatry*, *30*(12), 1983–1994. https://doi.org/10.1007/s00787-020-01667-5.

Ware, J., Kosinski, M., & Keller, S. (1998). *SF-12: how to score the SF-12 physical and mental health summary scales*

West, C., & Zimmerman, D. (1987). Doing Gender. *Gender & Society*, *7*(2), 125–151. https://doi.org/10.1177/0891243287001002002.

White, J., Connelly, G., Thompson, L., & Wilson, P. (2013). Assessing wellbeing at school entry using the Strengths and Difficulties Questionnaire: Professional perspectives. *Educational Research*, *55*(1), 8798. https://doi.org/10.1080/00131881.2013.767027.

Winterbottom, M., Smith, S., Hind, S., & Haggard, M. (2008). Understanding similarities and differences between parents' and teachers' construal of children's behaviour. *Educational Studies*, *34*(5), 483–510. https://doi.org/10.1080/03055690802288452.

Yoon, Y., Eisenstadt, M., Lereya, S.T., & Deighton, J. (2023). Gender difference in the change of adolescents' mental health and subjective wellbeing trajectories. *European Child & Adolescent Psychiatry*, *32*(9), 1569–1578. https://doi.org/10.1007/s00787-022-01961-4.