The effect of screen time on child behaviour: an instrumental variables approach

Jaimie Monk

Te Herenga Waka – Victoria University of Wellington

Abstract

Context: Children's screen use is a ubiquitous part of modern family life. However, nearly all empirical evidence of its effect on children's behaviour in the preschool years is associational – meaning that the effect of screen use on behaviour in this critical stage of development is relatively unknown.

This paper examines the effect of screen use on two-year-old children's behaviour using data from over 6,000 families in the Growing Up in New Zealand (GUiNZ) study.

Methods: This paper firstly explores associations between screen use and child behaviour using an ordinary least squares (OLS) regression. Secondly, to account for missing data, a Heckman correction is employed to address study attrition following multiple imputation of data from item non-response. Finally, an instrumental variable (IV) approach is adopted to isolate causality using two variables on family screen use rules as instruments.

Results: OLS results show a small association between higher levels of screen use and behaviour problems. However, a larger relationship is apparent when an IV approach is adopted.

Conclusion: These results suggest that associational estimates of the relationship between children's screen use and problem behaviour may underestimate the real effect. Therefore, the role of screen use in child behaviour problems may need increased consideration by policymakers.

1. Motivation

Screen use has become a significant part of modern childhood (Livingstone & Blum-Ross, 2020; Przybylski & Weinstein, 2019), and screens are now firmly embedded in the home environment of most New Zealand families (Colmar Brunton, 2015; New Zealand Ministry of Health, 2018). Due to their near-universal use, it has been argued that screen use should now be considered a fundamental part of the context in which child development occurs (Barr, 2019). However, firm conclusions on the effects of screen use on children, particularly preschool children, are inconclusive due to varying results, poor quality studies and little focus on determining causality. This uncertainty is compounded by suspected publication bias, with studies showing negative associations with screen use more likely to be published (Ophir et al., 2021).

Early intervention can improve outcome trajectories for young children and is known to be a cost-effective point of intervention. (Charach et al., 2017; Heckman, 2006). Consequently, a thorough understanding of whether screen use causes behavioural problems in very young children is important for both child wellbeing and to the development of effective and efficient policy interventions.

When it comes to the effects of screen use on child behaviour, the literature has grown rapidly in recent years. There has been some convergence confirming the harms of screen use on adolescent mental health, particularly for girls (e.g.Twenge & Farley, 2021), but little clear evidence on the effects of screen use on preschool children.

Arguably the principal current debate in the screen use literature is on the direction of causality – does screen use contribute to child behaviour problems? Or do children with problem behaviour get exposed to screens more frequently?

Until recently, this potential for bi-directional causality has been left relatively unaddressed. For example, in the meta-analysis of Nikkelen et al. (2014) addressing the effect of media use on ADHD-related behaviour, only three of the 45 empirical studies examined investigated the reverse relationship between ADHD-related behaviour and screen use. Nevertheless, more recent work has started to untangle the bi-directional relationship.

Several recent studies have tried to untangle this relationship explicitly. McDaniel and Radesky (2020) used structural equation modelling to find that higher preschool child externalising scores predicted higher parenting stress, which also predicted increases in child media use.[1] However, media use did not predict later externalising behaviour. However, results are limited by the small sample of primarily white parents and that the sample was of children of different ages.

Cliff et al. (2018) used data from the Longitudinal Study of Australian Children to find that lower television viewing and total media exposure at two years old was associated with higher self-regulation at four years of age. Lower self-regulation at four years was also associated with higher screen use at six years (although media exposure at four years was not associated with self-regulation at six years, and the size of associations was small).

However, the clearest evidence so far comes from Madigan et al. (2019) and Neville et al. (2021), who both use random-intercepts cross-lagged panel models that attempt to control for time-invariant differences between children. Madigan et al. assess the effect of screen time on children's achievement of developmental milestones (which includes personal/social skills) at 36 and 60 months. Higher levels of screen use were associated with subsequent poor performance in developmental tests, but the reverse association was not found.

Neville et al. modelled the relationship between screen use and internalising and externalising problems separately. In their sample of more than 10,000 Irish children, they found greater screen time at ages 3 and 5 were "directionally associated" with increased internalising problems at ages 5 and 7, respectively, but this was not the case for externalising problems. They also show both externalising and internalising problems being directionally associated with screen use in the other direction, but only for preschool children. Two big benefits of this study are that it provides some of the first reliable evidence that effects go in both directions and that the authors can control for the important within-child effects.

In summary, during the critical early preschool stage of child development the effect of screen use on child behaviour is relatively unknown due to a lack of quality evidence.

---

[1] Distinguishing problem behaviour between **externalising** and **internalising problems** is also common in the literature. While externalising behaviours tend to be displayed outwardly and are reflected by behaviour towards the physical environment (e.g. aggression, delinquency and hyperactivity), internalising disorders are more commonly directed inward and are indicative of a child's psychological and emotional state (e.g. anxiety, depressed behaviours, becoming withdrawn) (Liu et al., 2011). Internalising and externalising problems may often occur together (Achenbach et al., 2016).

2. Statistical methods

2.1. Data

Growing Up in New Zealand is a contemporary longitudinal study following 6,852 New Zealand children from birth to young adulthood. Parents were recruited from all expected births in the Auckland, Counties-Manukau and Waikato District Health Board regions of New Zealand between 25 April 2009 and 25 March 2010. In total, 6,822 women and 4,401 of their partners were recruited into the cohort, which is broadly generalisable to the New Zealand population in terms of ethnicity and markers of family socioeconomic status (Morton et al., 2015).[2]

Data for this study comes primarily from data collection wave two when the child is approximately two years old (83% of children are between 23-25 months old). However, some control variables are included from the antenatal (data collection wave 0) and nine-month waves (data collection wave 1). Appendix 1 outlines the source wave for each variable. Data for all three of these waves were collected using in-person interviews by trained interviewers. However, some background information was collected via phone call prior to the face-to-face interviews.

The focus of this study is mothers and their two-year-old children. While GUiNZ collected data from partners in earlier waves, there are significant demographic differences between the group of partnered mothers whose partner was in the study (N=3,852) and those whose partner was not (N=1,445).[3] It was therefore decided that the likely bias introduced by using partner data (and dropping those without partner data from the analysis) was going to be more detrimental to the accuracy of results than the benefit obtained by including partner-reported data in the analysis. A detailed comparison of the two groups is provided in Appendix 2. Of particular importance is the difference in mean child difficulties scores for those with partners in the study of 10.7 versus 12.73 for those without. Mean child screen use is also 1.24 hours per day for those with partners in the study versus 1.59 hours for those

[2] Of the 2,009 mothers who did not have a partner respond as part of the antenatal wave, 332 stated they did not have a partner they considered part of their family/whanau and 1,677 did but these partners were not included in the study.
[3] There are 336 mothers who had no partner (at the antenatal wave).

without. Consequently, both mothers with and without partners are therefore included in the sample for this study.

This study focuses on data at a single age (two years old). Two years of age is a unique stage of child development that involves a very sensitive and rapid period for socio-emotional development. Therefore, it is worth considering it in isolation from the rest of the preschool years. All families within the GUiNZ sample were included in this analysis.

Every Growing Up in New Zealand study participant had to provide informed consent. The Growing Up in New Zealand study had overall ethical approval from the Ministry of Health Northern B Regional Ethics Committee in New Zealand, and the Health and Disability Ethics Committee approves each subsequent data collection wave. Growing Up in New Zealand also complies with relevant University of Auckland guidelines for observational studies, the New Zealand Privacy Act and the New Zealand Health Research Council Guidelines.

2.2. Variables

Table 1 provides summary statistics for the key variables in this study. Other control variables are included in Appendix 3.

Table 1: Summary statistics for key variables

Variable	%/Mean	SD	Min	Max
Child difficulties score	11.455	5.146	0	31
Child screen use				
No screens	19.4			
0.1-1 hours	41.7			
1.1-2 hours	20.4			
2.1-3 hours	9.2			
3.1-4 hours	4.6			
4 hours+	4.6			
Control variables - antenatal wave				
Mother's ethnicity				
European	53.3			
Māori	13.9			
Pacific	14.6			
Asian	14.7			
Other	3.5			

Mother tertiary educated				
Yes	38.4			
No	61.6			
Mother's age	30.07	5.86	18	41
Control variables - 9 month wave				
Child gender				
Male	51.5			
Female	48.4			
Control variables - 2 year wave				
Income adequacy				
Not enough	10.4			
Just enough	32.9			
Enough	36.5			
More than enough	20.2			
Mother's overall stress	5.22	3.70	0	21
Extroversion	3.60	0.69	1.13	5
Agreeableness	3.97	0.50	1.89	5
Conscientiousness	3.99	0.57	1.22	5
Neuroticism	2.66	0.69	1	4.88
Openness	3.72	0.56	1.5	5
NZ Deprivation index score				
Low (1-3)	24.9			
Med (4-7)	36.8			
High (8-10)	35.8			
Mother paid job				
Yes	52.6			
No	47.4			
Partner status				
Has partner	90.0			
Does not have partner	10.0			

Source: Growing Up in New Zealand DCW0, DCW1, DCW2

The dependent variable of the child difficulties score has a mean of 11.45 and a standard deviation of 5.15. Screen use during the last weekday for most children was between 0.1-2 hours per day, with nearly 20% of children having no screen use and nearly 20% having more than 2 hours per day. For the key control variables, just over half of mothers gave European as their self-prioritised ethnicity, while 13.9% of mothers were Māori, 14.6% were Pacific and 14.7% were Asian. The mean age of mothers at the antenatal wave was 30.07 years old, while

38.4% had university education and 61.6% did not. Finally, 51.5% of the children were girls while 48.4 were boys.

The overall outcome measure in this study is child behaviour, measured using the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997), reported by mothers when children were two years old. The SDQ is a parent-rated 25-item scale that measures five aspects of child behaviour; emotional problems, peer relationship problems, hyperactivity/inattention, conduct problems and prosocial behaviour. The first four subscales are summed together to generate a total difficulties score (see Appendix 4 for a full list of the survey questions by subscale).[4] Prosocial behaviour provides the strengths score. However, this study focuses only on the difficulties score.

The SDQ was initially developed as a screening tool for pathological problems in child psychiatry and psychology. However, it is now widely used in large epidemiological studies and cohort development studies such as GUiNZ to measure child behaviour problems and is also used in the New Zealand B4 School Check administered to nearly all New Zealand preschool children before they start school. The SDQ has been used by over 4,000 international studies (Youth in Mind, 2022).[5]

In data collection, the "early-years" version of the SDQ was used (for ages 2-4), and it was asked of both mothers and partners with both fully answering the SDQ questionnaire. However, only mother data was used due to substantial missing partner data (see Appendix 1).

The screen use measure is composed of several questions (asked of the mother) on hours of screen use *during the last weekday,* from television, DVDs, computers and electronic gaming at the two-year-old wave. The responses from each type of screen use were added together to give a total screen use time.

The final total screen use measure had an unusual distribution. About 20% of children had no screen use, while those with positive screen use were clustered around half hour and full hour measurements - suggesting some measurement error. These two factors meant that screen use

[4] The SDQ involves mothers responding to 25 statements as to whether she considers them "Not true", "Somewhat true", or "Certainly true" of her child. Example questions include: "Often has temper tantrums or hot tempers", "Helpful if someone is hurt, upset or feeling ill", "Often argumentative with adults".

[5] However, there have been some criticisms of the cultural suitability of the SDQ in the New Zealand context. For a detailed overview see (2016).

had to be dealt with carefully when including it in regression models. To account for the high number of zeros and possible measurement error, the continuous measure was converted into an ordinal measure representing: no screen use, 0.1-1 hours screens, 1.1-2 hours, 2.1-3 hours, 3.1-4 hours and more than 4 hours for the OLS regressions. Measurement error is addressed when using instrumental variable (IV) analysis, so the continuous screen use measure is used for the IV estimates.

The model development strategy was to include a wide variety of control variables (due to concern regarding the potential for omitted variable bias). These variables were initially selected based on literature indicating associations with the relationships of interest. Correlations between the key independent variables and potential control variables were also examined using kernel density functions to compare how the variables were correlated across their distributions. Gelbach's decomposition was used to understand how the addition of each control variable affected the relationship between screen use and child behaviour.[6] VIF scores were used to determine if multicollinearity was a problem.[7] A full list of control variables and a more detailed definition of each variable is available in Appendix 1.

At two years of age the child is very strongly shaped by their direct environment, and in particular the relationship with their primary caregivers (specifically their mothers in the case of this study). Maternal demographic and personality variables included in the analysis are (self-reported) income adequacy, stress, education, self-prioritised ethnicity, age, employment and the Big Five Personality variables (extroversion, agreeableness, conscientiousness, neuroticism and openness), clinically significant post-natal depression symptoms (at 9 months) and prenatal stress. All these variables are empirically related to the key variables of interest in the literature and are shown to play an important role in the relationship between screen use and child behaviour.

Parenting has been shown by a number of studies to play an important role in child behavioural outcomes. There are a wide variety of parenting-related variables within the GUiNZ study, so measures were chosen carefully. The variables chosen were a measure of

[6] Gelbach (2016) has created a simple and effective way to account for this problem by using the omitted variable bias formula to construct a conditional decomposition to consider various covariates' (or groups of covariates') role in shifting base regressor's coefficients. The Gelbach decomposition was adopted to understand what effects the addition of covariates has on estimated coefficients for the key variables.

[7] The final OLS equation estimating child difficulties score had a mean VIF score of 1.40 with the highest individual VIF score from high hostile parenting with a score of 2.02

hostile parenting, one of positive parenting, a measure of protective parenting and whether a mother reads with her child. These measures capture broadly different aspects of parenting that are shown to be associated with both screen use and child behaviour. Maternal self-efficacy (a mother's self-perceived ability, competence and confidence as a mother) and a mother's level of personal support were also included due their effect on parenting behaviour. The number of hours a child was in childcare outside the home was initially included; however, it was ultimately dropped as VIF tests highlighted multicollinearity concerns with a mother's employment status.

Child related controls include child gender, a child's general health, their (parent-assessed) weight status, and the frequency of wakes in the night.

2.3. Methodology

For the first step in the analysis, a kernel density plot is used to compare the distributions of child difficulties scores for each subpopulation of child screen use category. Ordinary least squares (OLS) regression is then used to understand associations while controlling for the effects of other variables.

To address causality, two key areas were identified where bias might substantially affect results obtained using OLS regression. The first was from unmeasured factors affecting predictor and outcome variables (otherwise known as omitted variable bias) and the second was the likelihood of bi-directional effects (behaviour problems causing greater screen use as well as screen use causing the behaviour problems).

Adopting an instrumental variable approach was identified as the best way of controlling for both these sources of potential bias (subject to the identification of appropriate instruments). Instruments meeting standard tests for inclusion were found, so this approach was implemented. Using an instrumental variable approach also accounted for bias due to measurement error. In addition, a limitation of using a random-intercepts, cross-lagged modelling approach to determine causality (as other studied exploring this relationship tend to do) is that screen use may cause difficult behaviour in both the short and the long term, whereas the difficult behaviour is likely only to cause greater screen use in the short term. Using a cross-lagged model over several years is conceptually limited in this situation as it can only capture an association between variables contemporaneously. By contrast, an instrumental variable approach can make a valuable contribution to understanding causality.

Another area of concern was potential bias due to missing data. Data was missing due to item non-response (respondents not fully completing questionnaires) and from wave non-response (families dropping out of the study). It was decided to account for item non-response and wave non-response separately, with multiple imputation used for item non-response and a Heckman correction for wave non-response.

To investigate the relationship between child behaviour and screen use an initial model is specified:

(1): Difficulties = f(child screens use)

The control variables are then included in Equation (1) to give:

(2): Difficulties = f(child screen use, maternal controls, child controls, household controls)

Equations (1) and (2) are estimated using OLS regression. The robust option is adopted to give standard errors robust to the presence of heteroskedasticity.

This study uses a novel approach for dealing with potential bias due to missing data. Data missing from respondents skipping questions and those missing from study attrition are modelled separately, with multiple imputation applied in the first case and a Heckman correction applied in the second. In line with Graham (2012), this study will refer to missing data from those skipping questions as **item non-response** and data missing from families dropping out of the study as **wave non-response.**

The risk of bias from missing data depends on why data is missing. Using Rubin's (1976) terms, data can be **missing completely at random** (MCAR), **missing at random** (MAR), or **missing not at random** (MNAR). Data that is MCAR is missing due to factors that are entirely random and generally does not cause a problem. An example could be a weighing scale running out of batteries and several children missing a weight measurement by chance. However, MCAR is quite rare in practice and can be partially tested by testing to see if any variables in a dataset are related to missingness or through Little's missing not at random (MNAR) test (Little, 1988). If data is in fact MCAR, complete case analysis (often known as listwise deletion) can be used without concern for bias. Complete case analysis removes any individual with missing data from the sample. This removal is not a problem with MCAR data, although it can increase the size of standard errors due to a smaller sample size.

More often, missingness is due to factors observable within the sample; this is what Rubin refers to (somewhat misleadingly) as Missing at Random (MAR). Essentially, data is missing conditional on other variables in the model. For data that is MAR, complete case analysis will introduce bias into results, but the use of appropriate techniques to control for this bias using information from the available data can correct for it.

If missing data is systematically different from observed data and cannot be explained using variables in the dataset, then missing data is considered missing not at random (MNAR). MNAR data can occur either because the value itself determines the missingness (e.g., parents of children with compromised health who require more attentive care may be less likely to continue in a longitudinal study addressing the determinants of health) or because there is an underlying, unmeasurable reason for observations to be missing (such as a mother's motivation to participate in a study).

The first problem apparent with the GUiNZ data is that of the 6,852 families in the antenatal wave, 532 families are missing from the two-year wave – this is the wave response problem. Additionally, there are 497 children with data missing from one or more of the 25 SDQ questions in the two-year wave. Furthermore, after considering item non-response for all explanatory and control variables, the sample size decreases by a further 1,912 children if complete case analysis is used in the estimation of child difficulties scores (Equation 2).

Table 3: Sample size reduction due to missing data

Sample	**Sample size**
Full sample from antenatal wave	6,852
Two year wave	6,321
OLS equation for full model using complete case analysis	4,409

Source: Growing Up in New Zealand DCW0, DCW1, DCW2

The biggest item non-response problem for this study is from incomplete answers to the SDQ questions (data missing for the outcome variable). There were 497 children (7.8% of those in the 2-year wave) with incomplete scores. To account for this missingness, the mechanism of missingness needs to be well understood.

To understand the predictors of item non-response to difficulties scores a probit model is specified in order to estimate full response to the difficulties scores (see Appendix 5). Results show that Asian mothers and those from 'other' ethnicities have lower item-response to difficulties scores (85.6% and 86.4% respectively) compared to European mothers (94.1%), but Pacific and Māori mothers have rates in line with Europeans. In addition, a mothers age, New Zealand Deprivation Index score and whether the child was first born or not were all significantly associated with giving a full response.

These results suggest that the missing difficulties scores are not missing at random (NMAR) and therefore using complete case analysis is not appropriate. However, judging whether missing data is missing at random and can be explained by variables in the model or not missing at random requires a more nuanced understanding of the data.

The GUiNZ dataset has a variety of variables, including socioeconomic variables, personality variables, anxiety, depression, child variables, and mothers' self-efficacy. Using these in prediction models gives a reasonable degree of confidence that large causes of bias have been explained. Therefore, it is probably reasonable to assume that item non-response from independent variables is considered MAR due to the wide variety of variables used to predict missingness and the fact that item non-response is more random and is less affected than wave non-response by issues such as certain families being hard to contact. Nevertheless, a degree of caution should still be taken in interpreting results.

Multiple imputation (MI) was considered an appropriate way to deal with missing data due to item non-response. Multiple imputation deals with missing data by creating multiple sets of plausible values (to capture the uncertainty of the estimation of missing data), which are then pooled for analysis. MI works under the assumption that data is MAR, and other available data can therefore substitute for missingness. MI is best expressed as a process of three steps: (1) the imputation step, (2) the analysis step, and (3) the pooling step (Little & Rubin, 2002). The first step of multiple imputation involves replacing each missing value with an estimate. This estimate is obtained by using information from the other variables in the dataset, including specific predictor variables called auxiliary variables. This process is repeated to create multiple copies of the dataset with slightly different values for the missing data. The second step involves analysis using the model of interest on the multiple datasets, then finally, results are pooled together. In the case of this study, the imputation process was done using multivariate normal (MVN) distribution with 40 datasets.[8]

There are significant demographic differences between the families who completed the 2-year wave and those who did not (see Appendix 6). Data is clearly not missing completely at random. However, it is plausible to assume that parents with children with more difficult behaviour may find it harder to continue in a longitudinal study (meaning missingness is caused by the outcome variable itself and data is MNAR). To test this theory, two probit models are developed to predict participation in later waves of the study based on two-year difficulties scores, while controlling for income adequacy, mother's education, ethnicity, partner status, NZ Deprivation Index score, birth order, and mothers' stress and personality. Families with children with higher difficulties scores are found to be more likely to drop out of the study at later waves, suggesting that data missing due to wave non-response was likely to be MNAR (full results in Appendix 7). Therefore, using MI to account for potential bias is not appropriate as MI assumes data is either MCAR or MAR. However, a Heckman correction is well suited for situations where selection into the sample is not random.

The Heckman (1979) procedure controls for selection bias by treating the selection as an omitted variable problem. The Heckman correction is a two-step approach. The first step is a probit model predicting selection (the selection equation), in this case, whether a family selects into the two-year data collection wave or not. The second step uses ordinary least squares to estimate the ultimate dependent variable (using the outcome equation); in this case, the equation estimating child difficulties scores after controlling for selection bias.

The outcome equation is as follows:

$$y_i^* = x_{1i}'\beta_1 + \varepsilon_{1i} \qquad (i)$$

Where x_{1i}' denotes a vector of observable child and family characteristics and y_i^* denotes child *i*'s difficulties score in the two-year wave. The difficulties score is not observed for those children who are not in the two-year data collection wave (signified by the *).

A second equation is identified reflecting a binary outcome of yes or no to show whether a child's mother selects them into the two-year wave.

[8] 40 datasets was chosen in line with the rule of thumb which suggests the number of datasets should be at least equal to the percentage of incomplete cases (which is just over 30% in this case) (White et al., 2011).

$$h_i^* = x_{2i}'\beta_2 + \varepsilon_{2i} \qquad (ii)$$

ε_{1i} and ε_{2i} are assumed to follow a bivariate normal distribution according to the following rule:

$$y_i = y_i^*,\ h_i = 1 \ \ if\ h_i^* > 0 \qquad (iii)$$

$$y_i\ not\ observed,\ h_i = 0 \ \ if\ h_i^* \leq 0 \qquad (iv)$$

To factor in potential bias from selection, this process relies on the correlation between ε_{1i} and ε_{2i} and uses equation (ii) to create the inverse Mills ratio (IMR), λ_i, where

$$\lambda_i = \frac{\phi(x_{2i}'\beta_2)}{\Phi(x_{2i}'\beta_2)}$$

The IMR can be understood as the ratio of the probability density function [$\phi(x_{2i}'\beta_2)$] to the cumulative distribution function [$\Phi(x_{2i}'\beta_2)$] or, more generally, the likelihood of observing second stage data given first stage characteristics.

The IMR is then incorporated as an added regressor into equation (i), as shown in (v).

$$y_i = x_{1i}'\beta_1 + \sigma_{12}\lambda_i + \eta_i \qquad (v)$$

As well as including all variables from the structural equation (i) in the selection equation (ii), it is highly recommended that at least one variable should be included in the first stage that is not required in the second (Verbeek, 2017). This ensures that identification does not come from (untested) functional form assumptions (Little, 1985). These added variables are frequently referred to as 'exclusion restrictions' as it is assumed that they have no impact on the ultimate dependent variable, except indirectly through the IMR. Heckman selection models are sensitive to the choice of exclusion instruments (although many studies either do not include them or do not report them accurately (Lennox et al., 2012) (Wolfolds & Siegel, 2019)).

The construction of a Heckman correction using data from multiple waves of a longitudinal study provides some unique challenges. Firstly, the outcome equation of interest is constructed mainly of variables from the two-year wave. Therefore, including the same set of variables in the selection and outcome equations does not work as one cannot predict participation in the two-year wave based on observed data from the same two-year wave. Therefore, an alternative approach has been developed to exploit the benefits of using the

Heckman correction within these limitations. Firstly, λ_i is obtained by estimating the selection equation (ii) using only variables from the antenatal wave. λ_i, is then incorporated into outcome equation (v) which is restricted to the same variables from the antenatal wave used in the selection equation (except for the exclusion restriction variables).[9] As a robustness check, λ_i is incorporated into an unrestricted outcome equation that includes all two-year wave variables. The results from this unrestricted equation need to be interpreted with care as the selection and outcome equations have different variables, but combined with results from the first outcome equation, they indicate areas where bias could be a problem.[10] It should also be noted that the Heckman correction was performed on a single imputed dataset rather than on the complete 40 datasets.[11]

The child development process is complex; the factors that influence a child's development are numerous, frequently interrelated, and often difficult to measure accurately. While clearly illustrating the associations between key variables, an OLS regression model cannot capture everything influencing the relationships of interest. For example, variables may be missing from the models, leading to inaccurate results. For instance, how a mother was parented by her own parents is likely to influence both observable and unobservable characteristics and choices of the mother.

Omitted variables may cause the existing explanatory variables to be correlated with the error term – resulting in one form of **endogeneity** known as **omitted variable bias.** Results obtained in the presence of omitted variable bias do not reflect the true relationship between variables if the omitted variable is correlated with one of the explanatory variables.

In addition to the problem of omitted variable bias, difficult child behaviour could be causing higher screen use while concurrently screen use could be causing more problematic

[9] The exclusion restrictions chosen for this model were the District Health Board the mother was enrolled with (Auckland District Health Board, Counties Manukau District Health Board, or Waikato District Health Board) and the country in which the mother was born. Mothers in the Waikato region were more likely to be in the two-year wave, as were those born in New Zealand. These variables were insignificant when included independently in equation (6). For an exclusion restriction to be effective, it needs to create a correlation between the two error terms in equations (1) and (2). This relationship can be shown empirically by ensuring the significance of λ_i, which in this case has a p-value of 0.002 in the restricted model and 0.004 in the unrestricted model.

[11] This modification was because Stata's MI suite of commands did not support several of the commands needed to perform the Heckman correction manually.

behaviour. Such confounding of cause and effect can obscure the estimated relationship between variables and is another form of endogeneity known as **simultaneity bias**.

Dealing with omitted variable bias and simultaneity bias requires careful handling of endogeneity. Implementing a modelling approach that includes **instrumental variables** is a frequently adopted method to address both these potential problems.

An instrumental variable approach starts with a relationship between an outcome variable and one or more explanatory variables in which at least one explanatory variable is endogenous. An instrumental variable (often known as an "**instrument**" or the "**excluded instrument**") is a variable that does not appear in that relationship, and that is correlated with the endogenous explanatory variable but uncorrelated with the error term in the equation. Formally, we have:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

where Y is the outcome variable, X_1 is a vector of one or more endogenous explanatory variables (with coefficient vector β_1), X_2 is a vector of one or more exogenous explanatory variables (with coefficient vector β_2), and ε is the residual. We then require one or more exogenous excluded instruments, Z, whose number is at least as large as the number of endogenous variables. To be a suitable instrument, we require both: $Cov(Z, X_1) \neq 0$ and $Cov(Z, \varepsilon) = 0$.

In its two-stage least squares form, the endogenous explanatory variable(s) is first regressed on the instrument variable(s). The outcome variable is then regressed on the predicted variable(s) from the first stage to give an estimated relationship between the explanatory variable and the outcome of interest that is not subject to endogeneity bias. In modelling the relationship between screen use and child behaviour problems in this situation,

Each set of instrumental variable equations contains a consistent group of control variables (X_2) treated as exogenous in that equation. These variables are a mother's ethnicity, child gender, number of siblings, mother's disability status, and personality variables (extroversion, agreeableness, conscientiousness, neuroticism, openness). Ethnicity, child gender and the number of siblings are all plausibly exogenous, and a mother's disability status is also likely to be pre-determined and (mostly) beyond the mother's influence. The personality variables are less exogenous, but Cobb-Clark and Schurer (2012) indicate that personality traits are stable for adults, and because it is generally considered that these personality variables are largely pre-determined genetically and/or early in life (Shiner, 2015) (Caspi et al., 2003). Hence it is reasonable to treat them as exogenous.

Equations were estimated using the Stata command *ivreg2* (StataCorp, 2017b). The robust option was adopted to give standard errors robust to the presence of arbitrary heteroskedasticity.

The results presented in this study show estimates for the original dataset and a second set in which item non-response has been imputed using multiple imputation (MI). MI was applied using the same procedure outlined previously, creating 40 individual datasets.[12] Statistical tests for instrument suitability and the margins command are not available using the *cmdok* command. Therefore, a single imputed dataset was randomly selected to execute the tests and estimate margins. Appendix 9 demonstrates that the estimates for this single dataset and the 40 imputed datasets are very similar. This approach should, therefore, not materially affect instrument tests. The instrument variables have not had data imputed to avoid introducing endogeneity by using endogenous variables to predict missing instrument values.

An effective choice of instruments is at the centre of an accurate instrumental variable approach. Choosing the wrong instruments can lead to greater bias than what is obtained using OLS (Angrist & Krueger, 2001). The choice of instruments should be informed in the first instance by a theoretical understanding of the relationships between variables of interest. Secondly, to ensure instruments are correctly identified, three tests of instrument suitability are offered - tests for under-identification, weak instruments and over-identification. Each is

[12] The MI suite of commands in Stata does not directly support the ivreg2 command, so *cmdok: estimate ivreg2* was used for the MI data. The *cmdok* command allows estimation of certain commands not supported by Stata's MI suite of commands *(StataCorp, 2017a).*

available as part of Stata's *ivreg2* package.[13] Under-identification F-statistics, weak instrument F-statistics and Hansen J-statistics are reported, and all support the choice of instruments.[14]

In summary, this study first examines associations between screen use and child behaviour using an ordinary least squares (OLS) regression. To account for missing data, a Heckman correction is employed to address study attrition following multiple imputation of data from item non-response. Due to likely bias from omitted variables, bi-directional effects and measurement error, an instrumental variable (IV) approach is adopted to isolate causality using two variables on family screen use rules as instruments.

3. Results

3.1. Associational results

Figure 1 shows that the mean and variance of difficulties scores increase at higher levels of screen use.

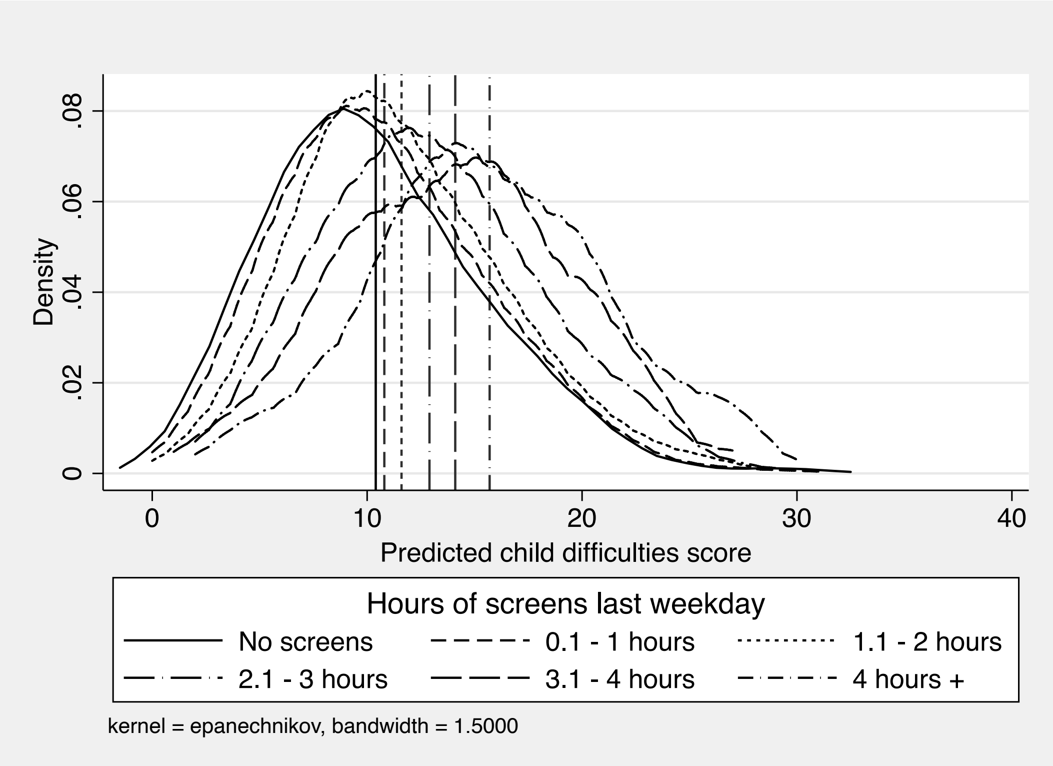
[13] The under-identification test tests the correlation between the instrument variables and the endogenous regressors. It is a Lagrange multiplier (LM) test of the null hypothesis that the instrument variables are appropriately correlated with the endogenous regressors (i.e., rejecting the null means the model is identified).

Weak instruments occur when there is a correlation between excluded instruments and endogenous regressors, but this correlation is weak. The weak instrument test reports a Kleibergen-Paap Wald F statistic; the Staiger and Stock (1997) rule of thumb is used where weak instruments are rejected if $F \leq 10$, although practitioners sometimes recommend a higher rejection cut-off as $F<20$.[13]

The over-identification test tests the null hypothesis that instruments are uncorrelated with the error term and that the excluded instruments are correctly excluded from the estimated equation. These tests require that the number of excluded instruments exceeds the number of endogenous variables in the equation and are reported using the Hansen J-statistic.

[14] Namely, the null hypothesis of under-identification has been rejected, the weak instrument test F-statistics were >10, and the null hypothesis of identification is not rejected for the overidentification test.

Figure 1: Child difficulties scores by screen use



Source: Growing Up in New Zealand DCW2

Table 1 presents results for Equation (1) and similarly shows that greater levels screen use are associated with steadily increasing levels of child difficulties scores.

Table 1: OLS results for Equation (1)

Child difficulties score	**(1)**
Screen use last weekday	
No screens	(base)
0.1 - 1 hours	0.394**
1.1 - 2 hours	1.168***
2.1 - 3 hours	2.559***
3.1 - 4 hours	3.681***
4 hours +	5.319***

Source: Growing Up in New Zealand DCW2

The results in Figure 1 and Table 1 do not however consider any other factors that could contribute to both screen use and child behaviour problems. Therefore, a regression model is needed in which these other factors can be controlled for. Control variables are therefore included in Equation (1) to give Equation (2).

The results, presented in Table 2 (and graphically in Figure 2), indicate that screen use at higher levels (of more than two hours per day) is noticeably associated with higher child difficulties scores. Full results with all control variables are presented in Appendix 10.

Looking at the results for the key control variables obtained using complete case analysis (column 1), mothers with a university education have children with approximately half a point lower difficulties scores than mothers without university education; Māori and Pacific mothers have children with difficulties scores that are 0.98 and 1.58 points higher than European mothers. Employed mothers and older mothers also have children with significantly lower difficulties scores. Girls have generally lower difficulties scores than do boys.

There is strong relationship between a mother's personality and her child's behaviour. Two factors are likely to be at play here, personality affecting parenting practices and styles, and a child inheriting (or copying) their parents' personality (Zwir et al., 2020). Higher difficulties scores are also significantly associated with higher levels of maternal stress.

Hostile parenting practices have the strongest relationship to child difficulties scores, with mothers who use these practices more frequently having children with difficulties scores that are 2.5 points higher on average than mothers who use them infrequently.

Finally, a mother's income adequacy, partner status or the household's New Zealand Deprivation Index scores are not associated with difficulties scores.

Table 2: OLS results for Equation (2)

Variable	Complete case analysis (1)	MI (2)	Imputed + Heckman (3)
Child screen use			
No screens	(base)	(base)	(base)
0 - 1 hours	0.057	0.138	0.106
1.1 - 2 hours	0.148	0.062	0.037
2.1 - 3 hours	0.780***	0.670***	0.719***
3.1 - 4 hours	0.885**	0.814**	0.811**
4 hours+	1.996***	1.878***	1.808***
Child gender			
Male	(base)	(base)	(base)
Female	-0.595***	-0.525***	-0.545***

(continued over page)

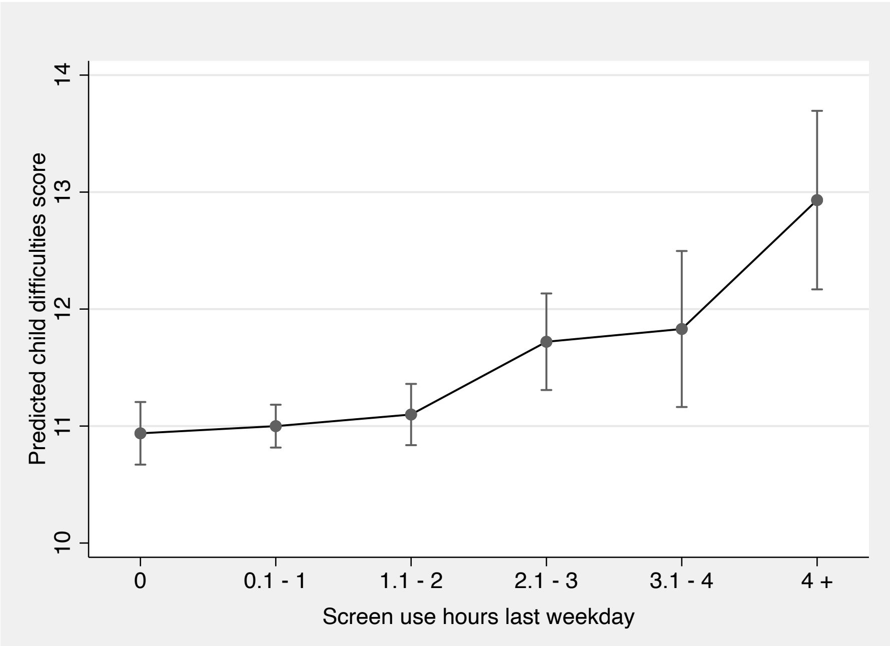
Income adequacy			
Not enough	0.139	0.187	0.289
Just enough	0.234	0.293	0.255*
Enough	0.066	0.133	0.125
More than enough	(base)	(base)	(base)
Mother's overall stress	0.129***	0.102***	0.119***
Extroversion	-0.132	-0.169*	-0.168*
Agreeableness	-0.444***	-0.409***	-0.429***
Conscientiousness	-0.936***	-0.604***	-0.580***
Neuroticism	0.403***	0.535***	0.511***
Openness	-0.548***	-0.524***	-0.504***
Mother's ethnicity			
European	(base)	(base)	(base)
Māori	0.978***	1.189***	0.798***
Pacific	1.582***	1.913***	1.263***
Asian	0.339	0.377**	-0.218
Other	0.162	0.162	-0.237
Mother tertiary educated			
Yes	-0.517***	-0.609***	-0.390***
No	(base)	(base)	(base)
Mother age	-0.078***	-0.071***	-0.057***
NZ Deprivation index score			
Low (1-3)	(base)	(base)	(base)
Med (4-7)	-0.029	0.025	0.028
High (8-10)	0.331*	0.378**	0.300**
Mother paid job			
Yes	-0.466***	-0.507***	-0.536***
No	(base)	(base)	(base)
Partner status			
Has partner	0.038	-0.103	-0.087
Does not have partner	(base)	(base)	(base)
Hostile parenting			
Low hostile parenting	(base)	(base)	(base)
Med hostile parenting	0.822***	0.813***	0.901***
High hostile parenting	2.560***	2.560***	2.644***
Other control variables	YES	YES	YES
Lambda			3.456**
Constant	20.837***	18.833***	18.476***

R -squared	0.373		0.373
Observations	4,409	6,091	6,068

*** p<0.01, ** p<0.05, * p<0.1

Source: Growing Up in New Zealand DCW0, DCW1, DCW2

Figure 2: OLS margins plot of child difficulties by screen use (95% CI)



Source: Growing Up in New Zealand DCW0, DCW1, DCW2

When it comes to the effect of bias from missing data, both item non-response and wave response will be addressed separately. Multiple imputation of item non-response caused estimates for the effects of screen use on child behaviour to decrease slightly for the coefficients at higher levels of screen use.

The biggest changes in estimates due to MI are for a mother's ethnicity, education and conscientiousness. For ethnicity, coefficient estimates increase for Māori, Pacific and Asian. The coefficient for maternal education increases in absolute value with MI but decreases for conscientiousness.

These results suggest that dropping cases using complete case analysis for item non-response may have caused bias by underestimating the effect size of ethnicity and overestimating the effect of stress and screen use on difficulties scores. There may have also been effects in relation to mothers with lower education and lower levels conscientiousness.

After accounting for wave non-response using the Heckman correction, the effect on screen use is insubstantial, except for a further decrease in the estimated coefficient for 4+ hours of screen use from 1.878 to 1.808. However, for ethnicity, the estimated coefficient for Māori mothers is now 0.798 (compared to 1.189 before the Heckman correction), and for Pacific mothers it is 1.263 (compared to 1.913). The coefficient for education is now -0.39 (compared to -0.609). Other estimated coefficients had only small adjustments once the Heckman correction was applied.

This leads to the big question left unanswered by this section, that of causality. While the relationships described above help understand the data, they can only reflect associations between the variables of interest. Additionally, omitted variable bias is a genuine concern when modelling complex relationships involving human behaviour. Consequently, an instrumental variable approach is needed to model the relationships between variables and determine causality with greater clarity.

3.2. IV Results

The OLS results in Table 2 indicate a significant relationship between a child's difficulties scores and their screen use for more than two hours of use (versus a base category of no screen use), but the effect size was not substantial. For instance, there is a predicted difficulties score of 10.94 for children with no screen use compared to 12.93 for those with more than 4 hours per day. When this relationship is modelled using an IV approach (using how often a mother keeps television rules and whether she has rules for the number of hours a child watches screens as instruments), the effect size becomes much larger with a coefficient of 1.197 (p-value <0.001). Screen use was measured in the OLS regressions as a categorical measure due to worries over measurement error from clustering around one and two hours of screen use. As an IV approach accounts for measurement error, the continuous screen use measure could be used.

Table 3 presents results for Equation (2) estimated using IV, followed by a comparison of the IV results with OLS results in Figure 3 and Table 4.

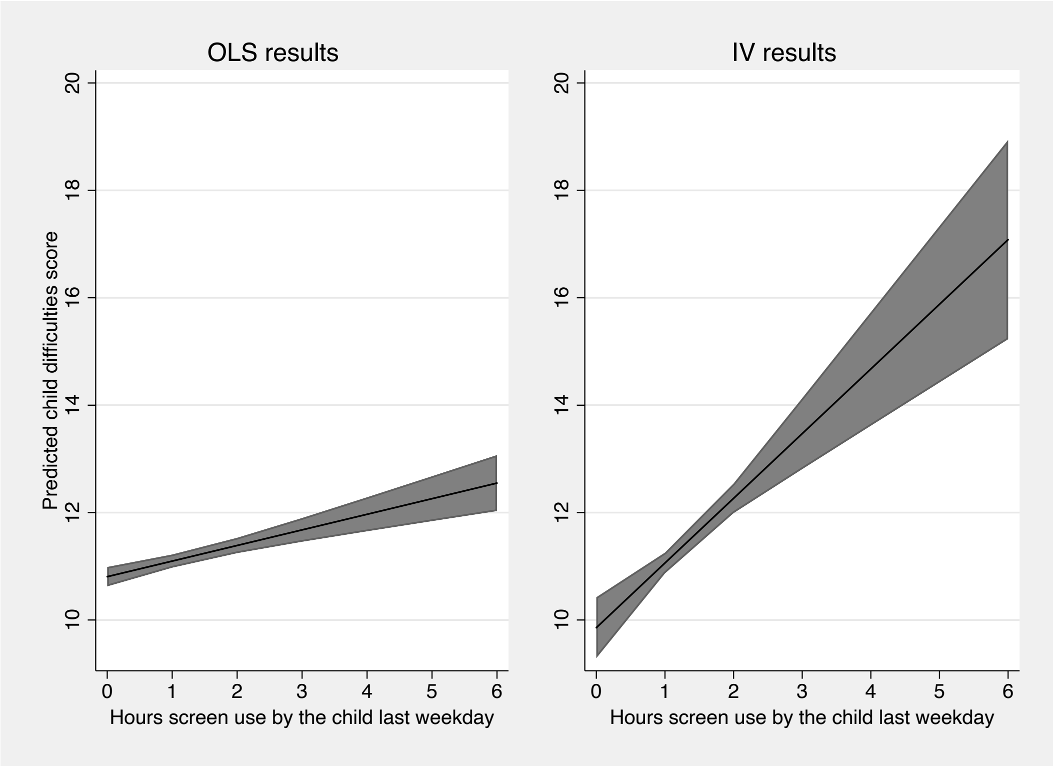
Table 3: IV results for Equation (2)

Dependent variable **Child difficulties**	(1) Original data	(2) MI
Screens last weekday	1.079***	1.197***
	(0.212)	(0.211)
Child's gender		
Boy	(base)	(base)
Girl	-0.499***	-0.546***
Ethnicity		
European	(base)	(base)
Māori	2.590***	2.543***
Pacific	3.125***	3.038***
Asian	0.421	0.370
Other ethnicity	0.509	0.511
Extroversion	-0.275***	-0.237**
Agreeableness	-0.729***	-0.728***
Conscientiousness	-1.079***	-1.020***
Neuroticism	1.289***	1.294***
Openness	-0.655***	-0.697***
Number of siblings		
Sole child	(base)	(base)
One sibling	0.016	0.016
Two siblings	-0.590***	-0.601***
Three + siblings	0.084	0.056
Mother's disability		
No	(base)	(base)
Yes	-0.369	-0.266
Constant	16.66***	16.35***
Imputations		40
N(observations)	4,928	5,524
Underidentification F-stat (p)	240.49	254.66
	(0.00)	(0.00)
Weak identification F-stat	46.87	41.97
Overidentification Hansen J-stat (p)	9.94	5.325
	(0.127)	(0.503)

*** p<0.01, ** p<0.05, * p<0.1

Source: Growing Up in New Zealand DCW0, DCW1, DCW2

Figure 3: Margins plot for difficulties scores by screen use for OLS and IV (95% CI)



Source: Growing Up in New Zealand DCW0, DCW1, DCW2

Table 4: The effect of screen use on child difficulties

	OLS			IV		
Hours	**Predicted difficulties**	95% CI		**Predicted difficulties**	95% CI	
0	**10.81**	10.62	10.99	**9.89**	9.29	10.42
1	**11.10**	10.97	11.22	**11.06**	10.87	11.25
2	**11.39**	11.25	11.53	**12.27**	11.99	12.54
3	**11.68**	11.46	11.90	**13.47**	12.81	14.13
4	**11.97**	11.65	12.28	**14.67**	13.62	15.73
5	**12.26**	11.84	12.68	**15.87**	14.42	17.33
6	**12.55**	12.03	13.07	**17.08**	15.23	18.93

Figure 3 and Table 4 show that children who have no screens have an estimated difficulties score of 9.89 and that this increases materially until children at 5 hours of screen use have an estimated score of 15.87, while for those at 6 hours per day, the score rises to 17.08. However, it should be noted that confidence intervals at higher levels of screen use become quite large due to smaller numbers of children in these groups. While acknowledging this caveat, these results indicate that avoiding high levels of screen use helps lower the likely number of behavioural problems for children.

In summary, OLS results show a small association between higher levels of screen use and behaviour problems. However, a larger relationship is apparent when an IV approach is adopted.

4. Limitations

Conclusions from this thesis may be limited because data obtained from partners has not been included (due to substantial missing data, which was found not to be missing at random and unable to be accounted for). While the partner variables did not significantly relate to the key variables of interest during scoping work for this project, a father still influences a child's life. Not including partners risks missing an essential part of the child development picture.

Figure 1 showed differences in variance between groups as well as differences in the mean. The regression methods used in this study (OLS, and IV) have not explicitly taken these differences of variance into account, so may have missed some distinctions in how variables relate to each other at different points in their distributions.

Social desirability bias should also be acknowledged in any longitudinal cohort study. Social desirability bias reflects the tendency of respondents to give socially desirable responses rather than choosing responses that are an accurate representation of their true feelings (Grimm, 2010). Most of the data used in this thesis comes from face-to-face interviews, where the possibility of socially desirable answers being provided is higher than for other methods such as online questionnaires. The most considerable risk of social desirability could be expected in the measure of screen use (if parents consider screen use a bad thing), in reporting child difficulties and in the measure of hostile parenting. In all these situations, social desirability bias would cause lower values to be reported, and therefore effect sizes would likely be underestimated. It is difficult to establish how big a problem this is, but if effect sizes are underestimated, correction for bias would tend to strengthen the conclusions of this study rather than detract from them.

Many of the general limitations of qualitative screen use research apply in this study. Firstly, because in GUiNZ, children's screen use is self-reported by parents, it is open to **response bias**. While parents may under-report screen use due to social desirability bias, there is also evidence to suggest parents over-report screen use due to measurement error (e.g. (Certain & Kahn, 2002), although more recent evidence has shown self-report can both under and over-report screen use compared with more objective measures of screen use (Radesky et al., 2020).

GUiNZ screen use data also covers only screen use in the home and cannot capture any screen use that occurs in formal care or while the child is under the care of other adults. Although the evidence suggests that screen use in New Zealand's formal early education settings is low (Gerritsen et al., 2016), there is little evidence on screen use in less formal settings.

The other limitation with screen use measurement is that GUiNZ only measures screen use on weekdays. Similar longitudinal studies have shown that screen use by children is substantially higher on weekend days than on weekdays (e.g. (e.g.,Australian Insititue of Family Studies, 2012; Growing up in Ireland, 2017). Hence using only weekday screen use will lead to an under-reporting of overall screen use. There is also potential for screen use patterns to be different on different weekdays (e.g., higher on a Friday). The day of screen use was not captured so could not be taken into account.

The quality of IV results is strongly affected by the choice of instruments. While care has been taken to establish a clear theoretical justification for the use of instruments, and they have passed the relevant statistical tests, there may still be unforeseen reasons why they are inappropriate. This caveat also applies to using the exclusion restriction instruments in the Heckman correction.

Another limitation is that the IV estimates involve more imprecision than the OLS estimates due to the instruments' strength. Nevertheless, once potential bias from bi-directional effects, omitted variables and measurement error are considered, these less precise IV estimates are judged to better reflect the true relationship between variables more accurately than the more precisely estimated (but likely inconsistent) OLS estimates.

5. Conclusions

Results indicate that child screen use could be an important source of behaviour problems for two-year old children in this study. To put these results in context, the SDQ was initially developed as a screening tool for child behaviour problems. The developers of the SDQ recommend a system of three categories in screening for problem behaviour. Children in the bottom eight deciles of observed difficulties scores are considered to have a normal score, those in the second-highest decile are considered in the borderline range, and those in the top decile are considered 'abnormal'. When these cut-offs are applied to the GUiNZ data, we get a cut off score of 16-18 in the borderline range and 19 and above for abnormal. The results from Table show that children who have more than 5 hours of screen use per day are on average in the borderline range, shifting to the abnormal range at the highest levels of screen use. These results also illustrate the importance of not relying on simple OLS regression models to understand the complex relationships involved in child development. Once omitted variable bias and bi-directional effects were controlled for, effect sizes increased. These results remained statistically significant despite larger standard errors – indicating that confounding factors in the OLS regressions were masking relationships between variables.

The biggest current discussion in the screen use and child behaviour literature centers on the direction of causality. Using IV to model this relationship has added the first contemporaneous, causal evidence required to isolate the effect of screen use on child behaviour from the effect of behaviour problems on screen use. It should be noted that this evidence relates to two-year-old children, and the relationship may well be different for children at different ages or in different contexts.

In addition, a novel technique of combining multiple imputation to account for item non-response, alongside a Heckman correction to account for wave dropout, was developed to examine the effect of missing data. Of specific interest was the influence of children with behaviour problems potentially dropping out of the GUiNZ study at greater rates. This technique did not substantially change estimates for the effect of screen use on child behaviour; however, estimates were changed for several control variables such as ethnicity and employment.

In summary, the first 1,000 days of a child's life are critical for a child's development and providing the right ingredients for healthy development helps determine a child's positive outcomes into adulthood. The near-universal use of screen time for children in these early years means understanding its role in healthy development is vital. This study provides causal evidence that exposure to screen use negatively effects children's behaviour at 2 years of age and therefore, the role of screen use in child behaviour problems may need increased consideration by policymakers.

Appendix 1: Variable description

Variable	Description	Wave sourced from
Key variables		
Child difficulties	From the Strengths and Difficulties Questionnaire. Mother-reported scale developed from 20 questions on child behaviour. Created by summing responses to four five-item subscales (emotional problems, peer relationship problems, hyperactivity/inattention and conduct problems). Scores can range from 0-40.	2 year
Income meets needs	Responses to "How well does your (and your partner's combined) total income meet your everyday needs for such things as accommodation, food, clothing and other necessities? Would you say you have not enough money, just enough money, enough money, or more than enough money?".	2 year
Mother's overall stress	The sum of responses to 8 questions on sources of stress: "Thinking about the time since your [child was/children were] nine months old, to what extent are the following sources of stress for you and your family." Ill or disabled family member, housing difficulties, balancing work and family life, money problems, family members not getting on, another child's behaviour, parenting the study child, who does household chores. Overall stress scale is from 0-30.	2 year
Total screens	Mother reported number of hours *last weekday* spent watching television, dvds, or using a laptop, children's computer system or electronic gaming system.	2 year
Control variables		
Ethnicity	Mother self-identified and self-prioritised ethnicity. Categorised as New Zealand European, Maori, Pacific, Asian and other.	Antenatal
Mother tertiary educated	Binary variable. Mother has bachelor's degree/higher degree or not.	Antenatal
Mother age	Mother's age in years.	Antenatal
Prenatal perceived stress	Perceived stress scale. Scale from 0-40. Derived from (Cohen et al., 1983)	Antenatal
Mother's general health	Answer to the question "Thinking about before you became pregnant, in general how would you say your health was?" Response options are poor, fair, good, very good and excellent.	Antenatal
Child gender	Male or female	9 month
Child health	Mother reported binary variable. In response to "In general, how would you say baby's current health is?" Categorised into Excellent/very good and Poor/fair/good.	9 month

Maternal self-efficacy	Extract from the Pridham scale. (Pridham & Chang, 1989). Nine items from the original Pridham scales plus two extra questions about overall parenting confidence and mother-child closeness. Scores range from 6-66.	9 month
Personal support	Parenting Social Support Scale (Dunst et al. (1984)). A measure derived from 10 questions asking about support from a mothers partner, wider family and support services (e.g. doctors). Individual questions are scored from 1 (not available) to 6 (extremely helpful). Overall scores range from 12-60.	9 month
Clinically significant PND symptoms	Derived from 10 item Edinburgh post-natal depression scale (Cox et al., 1987). Original scores from 0-30. Clinically significant cut off point of 13 or more.	9 month
Number of siblings	The number of siblings a child has living with them at home.	16 month
Extroversion	Derived from the Big Five Inventory – Adolescent version (chosen due to simpler text than adult version)(John & Srivastava, 1999). Scale from 0-5.	2 year
Agreeableness	Derived from the Big Five Inventory – Adolescent version (chosen due to simpler text than adult version)(John & Srivastava, 1999). Scale from 0-5.	2 year
Conscientiousness	Derived from the Big Five Inventory – Adolescent version (chosen due to simpler text than adult version)(John & Srivastava, 1999). Scale from 0-5.	2 year
Neuroticism	Derived from the Big Five Inventory – Adolescent version (chosen due to simpler text than adult version)(John & Srivastava, 1999). Scale from 0-5.	2 year
Openness	Derived from the Big Five Inventory – Adolescent version (chosen due to simpler text than adult version)(John & Srivastava, 1999). Scale from 0-5.	2 year
NZ Deprivation Index	Categorised New Zealand Deprivation Index score from 2006. Low = 1-3, Medium = 4-7, High = 8-10.	2 year
Mother paid job	Answer to “Do you have a paid job at the current time?”	2 year
Government benefit	Mother receives government benefit (excluding Working for Families). These include: unemployment benefit, sickness benefit, NZ Superannuation, domestic purposes benefit, invalids benefit, student allowance, regular ACC payments or “other government benefits”.	2 year
Partner status	Answer to “Do you have a current partner?”	2 year
Moves since antenatal wave	Number of times mother has moved house since the antenatal wave.	2 year
Home ownership	Home ownership classified by personal ownership, private rental, public rental or "other".	2 year

Wakes in night	The number of times the child wakes in the night, on average	2 year
Positive parenting	Time Spent With Child Scale (Davies et al., 2002). Sum of 12 questions of whether parent engages positively with child.	2 year
Hostile parenting	Sum of response to four questions "During the past 4 weeks how often did you…" get angry at him/her, criticize his/her ideas, shout at him/her, argue when disagree with him/her. Categorised into low/medium/high.	2 year
Protective parenting	Sum of 4 questions on protective parenting: "How often do you try to protect child from life's difficulties?" "How often do you put child's needs and wants before your own?" "How often does leaving child with other people upset you no matter how well you know them?" "How often do you let child take a risk if there is no major threat to [his/her] safety?" Categorised into low/medium/high.	2 year
Mother reads with child	"How often do you read books with your child?" Low = never/seldom/several times a week Medium= Daily High = Several times a day	2 year

Appendix 2: Comparing partnered mothers with and without partners participating in the study

	Partner in study (%)	Partner not in study (%)	Test of difference between groups
Child difficulties score			p<0.0001
(Mean)	10.70	12.73	
Income meets needs			p<0.0001
Not enough	8.17	13.59	
Just enough	30.98	34.89	
Enough	37.34	35.84	
More than enough	23.51	15.69	
Maternal stress			p<0.0001
(Mean)	5.00	5.51	
Child screen use (hours)			p<0.0001
(Mean)	1.24	1.59	
Mother university educated			p<0.0001
Yes	46.54	25.57	
No	53.46	74.43	
Ethnicity			p<0.0001
European	61.6	38.66	
Maori	9.93	19.93	
Pacific	9.50	23.75	
Asian	15.16	14.86	
Other	3.81	2.80	
Mothers' age			p<0.0001
(Mean)	30.75	28.92	
NZ Deprivation score			p<0.0001
Low (1-3)	28.62	19.25	
Medium (4-7)	39.31	32.66	
High (8-10)	32.07	48.09	
Number of mothers	3,852	1,445	

Appendix 1.1 shows the differences in key variables between mothers with partners who had those partners included in the study and those that did not. Test for differences between the groups were a chi-squared test for the categorical variables and a two-sample t-test for the difference in means for the continuous variables.

Appendix 3: Extra summary statistics

Variable	%/Mean	SD	Min	Max
Prenatal perceived stress	13.23	6.43	0	40
Child health				
Excellent/very good	85.7			
Poor/fair/good	14.3			
Personal support	32.79	7.2	12	60
Maternal self-efficacy	59.93	4.63	32	66
Clinically sig. PND symptoms				
Yes	8.1			
No	91.8			
Maternal general health				
Excellent	20.5			
Very good	35.3			
Good	34.0			
Fair	8.01			
Poor	2.2			
Number of siblings at home				
No siblings	39.8			
One sibling	35.2			
Two siblings	15.2			
Three+ siblings	9.7			
Parent assessed weight				
Underweight	10.0			
Normal weight	82.7			
Overweight	7.4			
Child wakes in night				
Sleeps through	50.7			
Wakes once	32.4			
Wakes 2+ times	16.8			
Income from govt. benefit				
Yes	25.1			
No	74.9			
Moves since antenatal wave				
No moves	55.4			
One move	27.5			
Two + moves	17.1			
Home ownership				
Owns home	52.9			

Private rental	38.9
Public rental	6.58
Other	1.62
Positive parenting	
Low positive parenting	19.6
High positive parenting	80.4
Hostile parenting	
Low hostile parenting	21.6
Med hostile parenting	38.5
High hostile parenting	39.9
Protective parenting	
Low protective parenting	29.6
Med protective parenting	32.1
High protective parenting	38.3
Mother reads with child	
Low reading	33.3
Med reading	29.2
High reading	37.5

Appendix 4: SDQ questions

Internalising problems
Emotional problems
Often complains of headaches, stomach-aches, or sickness.
Many worries, often seems worried.
Often unhappy, down-hearted or tearful.
Nervous or clingy in new situations, easily loses confidence.
Many fears, easily scared
Peer relationship problems
Rather solitary, tends to play alone.
Has at least one good friend.
Generally liked by other children.
Picked on or bullied by other children.
Gets on better with adults than with other children.
Externalising problems
Conduct problems
Often has temper tantrums or hot tempers
Generally obedient, usually does what adults request
Often fights with other children or bullies them
Often argumentative with adults
Can be spiteful to others
Hyperactivity/inattention
Restless, overactive, cannot stay still for long
Constantly fidgeting or squirming
Easily distracted, concentration wanders
Can stop and think things out before acting
Sees tasks through to the end, good attention span

Each questions had the response options of “Not true”, “Somewhat true” and “Always true”. Note that positive items were reverse coded.

Appendix 5: Predicting item response

Full response to difficulties score	(1)
Antenatal wave variables	
Mother's prioritised ethnicity	
European	(base)
Maori	0.085
Pacific	-0.095
Asian	-0.503***
Other	-0.470***
Mother university education	
No	(base)
Yes	-0.059
Mother age (years)	0.015***
Parity	
First born	(base)
Subsequent birth	0.199***
9 month wave variables	
Maternal evaluation/self-efficacy	-0.024
2 year wave variables	
Income meets needs	
Not enough	-0.037
Just enough	-0.093
Enough	-0.027
More than enough	(base)
NZ Deprivation index score	
Low (1-3)	(base)
Medium (4-7)	-0.128*
High (8-10)	-0.209***
Paid job	
No	(base)
Yes	0.090*
Overall stress	0.0028
Extroversion	0.065

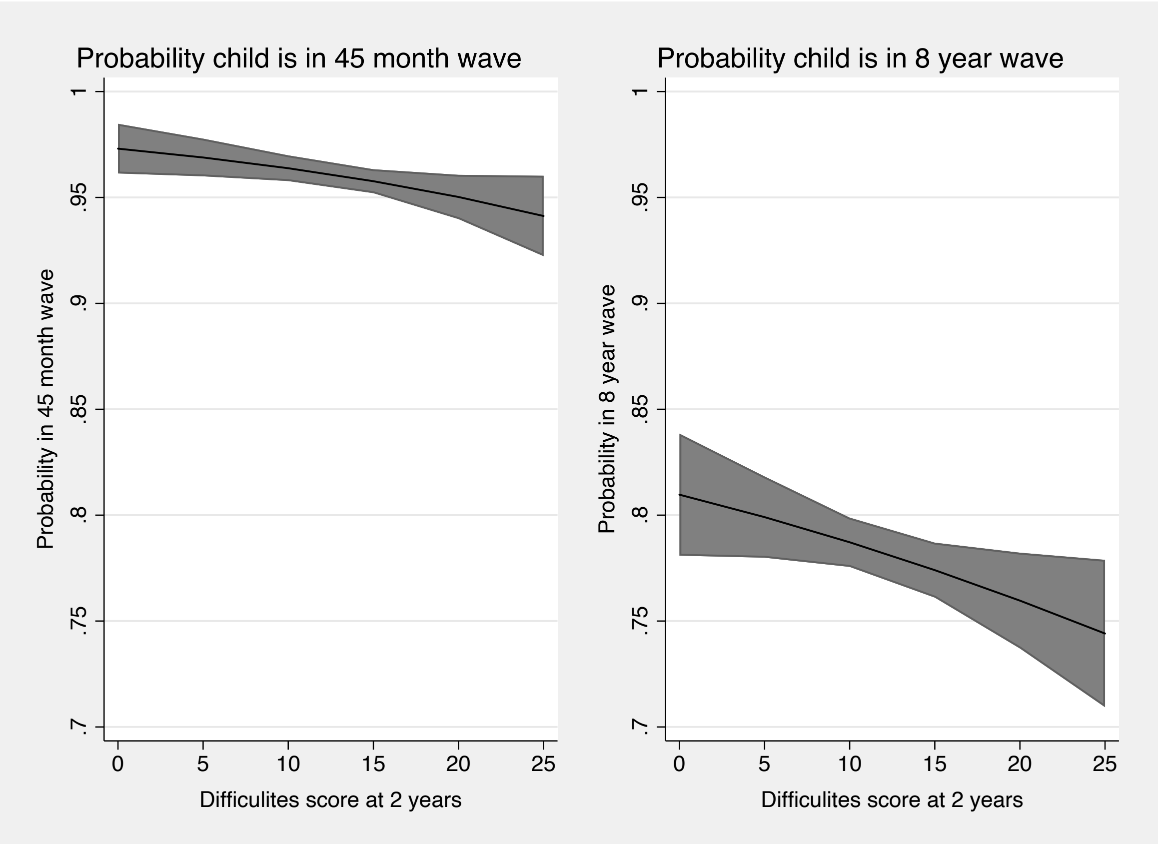
Agreeableness	-0.052
Conscientiousness	0.051
Neuroticism	-0.038
Openness	0.085*
Constant	1.484***
Pseudo R-squared	0.0431
Observations	5,861

Appendix 6: Differences between mothers who completed and dropped out of the 2-year wave

	Completed two year wave (%)	Dropped out of two year wave (%)	Test of difference between groups
Household income			p<0.0001
<20k	3.73	11.21	
20k-30k	5.2	10.91	
30k-50k	13.24	24.85	
50k-70k	16.32	18.48	
70k-100k	23.55	17.88	
100k-150k	22.87	12.42	
150k+	15.08	4.24	
Mother university educated			p<0.0001
Yes	39.99	19.4	
No	60.01	80.6	
Ethnicity			p<0.0001
European	56.14	20.52	
Maori	13.32	20.71	
Pacific	13.24	30.41	
Asian	13.78	25	
Other	3.51	3.36	
Mothers' age			p<0.0001
(Mean)	30.27	27.75	
NZ Deprivation score			p<0.0001
Low (1-3)	25.82	14.53	
Medium (4-7)	37.62	24.21	
High (8-10)	36.56	61.27	

Appendix 7: Predicting wave drop out

Probability of child remaining in future waves by two-year difficulties score (95%CI)



Appendix 8: Heckman selection results for difficulties scores (restricted model)

Child difficulties	(1) MI data	(2) Heckman estimation	(3) Heckman selection
Age	-0.133***	-0.111***	0.0219***
Ethnicity			
European	(base)	(base)	(base)
Maori	2.311***	1.842***	-0.465***
Pacific	3.083***	2.202***	-0.483***
Asian	1.487***	0.582	-0.384***
Other ethnicity	0.671*	0.288	-0.271*
Mother university educated			
Yes	-1.07***	-0.865***	0.231***
No	(base)	(base)	(base)
NZ Deprivation score			
Low (1-3)	(base)	(base)	(base)
Med (4-7)	0.035	0.094	-0.055
High (8-10)	0.652***	0.509**	-0.095
Child gender			
Male	(base)	(base)	(base)
Female	-0.687***	-0.711***	-0.023
Perceived stress	0.150***	0.147***	0.0025
No. moves in last 5 years	0.042*	0.0066	-0.0339***
District Health Board			
Auckland DHB			(base)
Counties Manukau DHB			-0.0057
Waikato DHB			0.257***
Mother's place of birth			
New Zealand			(base)
Australia			-0.070
Other Oceania			-0.267***
Asia			-0.398***
Other			-0.187
Lambda		5.590***	
Constant	12.995	11.993***	1.24***
Observations	6,268	6,268	6,268

Appendix 9: Comparing a single and 40 imputations

	(1)	(2)
Dependent variable **Child difficulties**	MI single imputation	MI 40 imputations
Screens last weekday	1.204***	1.197***
	(0.205)	(0.211)
Child's gender		
Boy	(base)	(base)
Girl	-0.587***	-0.546***
Ethnicity		
European	(base)	(base)
Maori	2.584***	2.543***
Pacific	3.031***	3.038***
Asian	0.256	0.370
Other ethnicity	0.410	0.511
Extroversion	-0.257***	-0.237**
Agreeableness	-0.687***	-0.728***
Conscientiousness	-1.064***	-1.020***
Neuroticism	1.275***	1.294***
Openness	-0.699***	-0.697***
Number of siblings		
Sole child	(base)	(base)
One sibling	0.046	0.016
Two siblings	-0.651***	-0.601***
Three + siblings	0.058	0.056
Mother's disability		
No	(base)	(base)
Yes	-0.409	-0.266
Constant	16.54***	16.35***
Imputations	1	40
N(observations)	5,524	5,524

*** p<0.01, ** p<0.05, * p<0.1

Appendix 10: Heckman selection results for difficulties scores (unrestricted model)

	Complete case analysis	MI	Imputed + Heckman
Antenatal wave variables	(1)	(2)	(3)
Mother's ethnicity			
European	(base)	(base)	(base)
Māori	0.978***	1.189***	0.798***
Pacific	1.582***	1.913***	1.263***
Asian	0.339	0.377**	-0.218
Other	0.162	0.162	-0.237
Mother tertiary educated			
Yes	-0.517***	-0.609***	-0.390***
No	(base)	(base)	(base)
Mother age	-0.078***	-0.071***	-0.057***
Prenatal perceived stress	0.027**	0.030**	0.027**
9 month wave variables			
Child gender			
Male	(base)	(base)	(base)
Female	-0.595***	-0.525***	-0.545***
Child health			
Excellent/very good	-0.924***	-0.821***	-0.868***
Poor/fair/good	(base)	(base)	(base)
Maternal self-efficacy	-0.038**	-0.028**	-0.033**
Clinically sig. PND symptoms			
Yes	(base)	(base)	(base)
No	-0.021	-0.078	-0.066
Personal support	-0.0008	-0.009	-0.005
2 year wave variables			
Extroversion	-0.132	-0.169*	-0.168*
Agreeableness	-0.444***	-0.409***	-0.429***
Conscientiousness	-0.936***	-0.604***	-0.580***
Neuroticism	0.403***	0.535***	0.511***
Openness	-0.548***	-0.524***	-0.504***
Number of siblings			
No siblings	(base)	(base)	(base)
One sibling	-0.155	-0.234*	-0.232*
Two siblings	-0.525***	-0.548***	-0.558***

Three+ siblings	0.194	0.004	0.006
	Complete case analysis	MI	Imputed + Heckman
Child wakes in night			
Sleeps through	(base)	(base)	(base)
Wakes once	0.347**	0.275**	0.296**
Wakes 2+ times	0.592***	0.375**	0.398**
Child weight (parent assessed)			
Underweight	0.539**	0.537***	0.538***
Normal weight	(base)	(base)	(base)
Overweight	0.498**	0.424**	0.394**
NZ Deprivation index score			
Low (1-3)	(base)	(base)	(base)
Med (4-7)	-0.029	0.025	0.028
High (8-10)	0.331*	0.378**	0.300**
Mother paid job			
Yes	-0.466***	-0.507***	-0.536***
No	(base)	(base)	(base)
Income from govt. benefit			
Yes	0.380**	0.419***	0.401***
No	(base)	(base)	(base)
Partner status			
Has partner	0.038	-0.103	-0.087
Does not have partner	(base)	(base)	(base)
Moves since antenatal wave			
No moves	(base)	(base)	(base)
One move	0.006	-0.080	-0.119
Two + moves	0.253	0.190	0.161
Home ownership			
Owns home	(base)	(base)	(base)
Private rental	-0.132	-0.072	-0.017
Public rental	0.560	0.416	0.558**
Other	0.378	0.245	0.191
Positive parenting			
Low positive parenting	0.628***	0.623***	0.635***
High positive parenting	(base)	(base)	(base)

(continued over page)

	Complete case analysis	MI	Imputed + Heckman
Hostile parenting			
Low hostile parenting	(base)	(base)	(base)
Med hostile parenting	0.822***	0.813***	0.901***
High hostile parenting	2.560***	2.560***	2.644***
Protective parenting			
Low protective parenting	(base)	(base)	(base)
Med protective parenting	0.139	0.162	0.108
High protective parenting	0.420***	0.372**	0.296***
Mother reads with child			
Low reading	(base)	(base)	(base)
Med reading	-0.508***	-0.522***	-0.575***
High reading	-0.640***	-0.700***	-0.851***
Lambda			3.456**
Constant	20.837***	18.833***	18.476***
R -squared	0.373		0.373
Observations	4,409	6,091	6,068

*** p<0.01, ** p<0.05, * p<0.1

Bibliography

Achenbach, T. M., Ivanova, M. Y., Rescorla, L. A., Turner, L. V., & Althoff, R. R. (2016). Internalizing/externalizing problems: Review and recommendations for clinical and research applications. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(8), 647-656.

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives, 15*(4), 69-85.

Australian Insititue of Family Studies. (2012). *The Longitudinal Study of Australian Children: Annual statistical report 2011*. Retrieved from https://growingupinaustralia.gov.au/sites/default/files/asr2011.pdf

Barr, R. (2019). Growing up in the digital age: early learning and family media ecology. *Current Directions in Psychological Science, 28*(4), 341-346.

Caspi, A., Harrington, H., Milne, B., Amell, J. W., Theodore, R. F., & Moffitt, T. E. (2003). Children's behavioral styles at age 3 are linked to their adult personality traits at age 26. *Journal of personality, 71*(4), 495-514.

Certain, L. K., & Kahn, R. S. (2002). Prevalence, correlates, and trajectory of television viewing among infants and toddlers. *Pediatrics, 109*(4), 634-642.

Charach, A., Bélanger, S. A., McLennan, J. D., & Nixon, M. K. (2017). Screening for disruptive behaviour problems in preschool children in primary health care settings. *Paediatrics & child health, 22*(8), 478-484.

Cliff, D. P., Howard, S. J., Radesky, J. S., McNeill, J., & Vella, S. A. (2018). Early childhood media exposure and self-regulation: bidirectional longitudinal associations. *Academic pediatrics, 18*(7), 813-819.

Cobb-Clark, D. A., & Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters, 115*(1), 11-15.

Colmar Brunton. (2015). *Children's media use study: How children engage with media today*. Retrieved from Wellington:

Gelbach, J. B. (2016). When do covariates matter? And which ones, and how much? *Journal of Labor Economics, 34*(2), 509-543.

Gerritsen, S., Morton, S. M., & Wall, C. R. (2016). Physical activity and screen use policy and practices in childcare: results from a survey of early childhood education services in New Zealand. *Australian and New Zealand journal of public health, 40*(4), 319-325.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry, 38*(5), 581-586.

Graham, J. W. (2012). *Missing data: Analysis and design*. New York: Springer New York.

Grimm, P. (2010). Social desirability bias. In W. Kamakura (Ed.), *Wiley international encyclopedia of marketing* (Vol. 2): John Wiley & Sons.

Growing up in Ireland. (2017). *Growing up in Ireland: Key findings at 7/8 years - socio-emotional development and play*. Retrieved from https://www.growingup.ie/pubs/GUI-KF-3.pdf

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(5782), 1900-1902.

Kersten, P., Dudley, M., Nayar, S., Elder, H., Robertson, H., Tauroa, R., & McPherson, K. M. (2016). Cross-cultural acceptability and utility of the strengths and difficulties questionnaire: views of families. *BMC psychiatry, 16*(1), 1-9.

Lennox, C. S., Francis, J. R., & Wang, Z. (2012). Selection models in accounting research. *The accounting review, 87*(2), 589-616.

Little, R. J. (1985). A note about models for selectivity bias. *Econometrica: Journal of the econometric society, 53*(6), 1469-1474.
Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association, 83*(404), 1198-1202.
Little, R. J., & Rubin, D. B. (2002). Bayes and multiple imputation. In R. J. Little & D. B. Rubin (Eds.), *Statistical analysis with missing data* (pp. 200-220). Hoboken, NJ: John Wiley & Sons.
Liu, J., Chen, X., & Lewis, G. (2011). Childhood internalizing behaviour: analysis and implications. *Journal of psychiatric and mental health nursing, 18*(10), 884-894.
Livingstone, S., & Blum-Ross, A. (2020). *Parenting for a digital future: How hopes and fears about technology shape children's lives*. New York: Oxford University Press.
Madigan, S., Browne, D., Racine, N., Mori, C., & Tough, S. (2019). Association between screen time and children's performance on a developmental screening test. *JAMA pediatrics, 173*(3), 244-250.
McDaniel, B. T., & Radesky, J. S. (2020). Longitudinal associations between early childhood externalizing behavior, parenting stress, and child media use. *Cyberpsychology, Behavior, and Social Networking, 23*(6), 384-391.
Morton, S. M., Ramke, J., Kinloch, J., Grant, C. C., Carr, P. A., Leeson, H., . . . Robinson, E. (2015). Growing Up in New Zealand cohort alignment with all New Zealand births. *Australian and New Zealand journal of public health, 39*(1), 82-87.
Neville, R. D., McArthur, B. A., Eirich, R., Lakes, K. D., & Madigan, S. (2021). Bidirectional associations between screen time and children's externalizing and internalizing behaviors. *Journal of child psychology and psychiatry, 62*(12), 1475-1484.
New Zealand Ministry of Health. (2018). *Annual update of key results 2017/18: New Zealand Health Survey*. Retrieved from Wellington: https://www.health.govt.nz/publication/annual-update-key-results-2017-18-new-zealand-health-survey
Nikkelen, S. W., Valkenburg, P. M., Huizinga, M., & Bushman, B. J. (2014). Media use and ADHD-related behaviors in children and adolescents: A meta-analysis. *Developmental psychology, 50*(9), 2228.
Ophir, Y., Rosenberg, H., & Tikochinski, R. (2021). What are the psychological impacts of children's screen use? A critical review and meta-analysis of the literature underlying the World Health Organization guidelines. *Computers in Human Behavior, 124*, 106925.
Przybylski, A. K., & Weinstein, N. (2019). Digital Screen Time Limits and Young Children's Psychological Well-Being: Evidence From a Population-Based Study. *Child development, 90*(1), e56-e65.
Radesky, J. S., Weeks, H. M., Ball, R., Schaller, A., Yeo, S., Durnez, J., . . . Coyne, S. (2020). Young children's use of smartphones and tablets. *Pediatrics, 146*(1), e20193518.
Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592.
Shiner, R. L. (2015). The development of temperament and personality traits in childhood and adolescence. In M. Mikulincer, P. Shaver, L. Cooper, & R. Larsen (Eds.), *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences.* (pp. 85-105). Washington: American Psychological Association.
Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557-586.
StataCorp. (2017a). Stata Multiple-Imputation Reference Manual. College Station, TX: StataCorp LLC.
StataCorp. (2017b). Stata Statistical Software (Version Release 15). College Station, TX: StataCorp LLC.

Twenge, J. M., & Farley, E. (2021). Not all screen time is created equal: associations with mental health vary by activity and gender. *Social psychiatry and psychiatric epidemiology, 56*(2), 207-217.
Verbeek, M. (2017). *A guide to modern econometrics* (5 ed.). Hoboken, NJ: John Wiley & Sons.
White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine, 30*(4), 377-399.
Wolfolds, S. E., & Siegel, J. (2019). Misaccounting for endogeneity: The peril of relying on the Heckman two-step method without a valid instrument. *Strategic Management Journal, 40*(3), 432-462.
Youth in Mind. (2022). Uses of the SDQ. Retrieved from https://youthinmind.com/products-and-services/sdq/
Zwir, I., Arnedo, J., Del-Val, C., Pulkki-Råback, L., Konte, B., Yang, S. S., . . . Garcia, D. (2020). Uncovering the complex genetics of human character. *Molecular psychiatry, 25*(10), 2295-2312.