Developmental complexity, structural simplicity: a longitudinal, multi-method investigation of internalising and externalising symptoms in young people

Praveetha Patalay

UCL

Thesis submitted for the degree of Doctor of Philosophy (PhD)

2015

I, Praveetha Patalay confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This research involves a series of studies investigating various aspects of the development of mental disorder during late childhood and early adolescence (8-14 years) in a large nationwide longitudinal study in England. The first two studies examine the clinical validity and survey format equivalence of a child self-report measure of mental health, the Me and My School questionnaire, measuring symptoms in the two key domains of child psychopathology: internalising and externalising.

The next two studies investigate the complexity of individuals' symptom development by examining the developmental trajectories of internalising and externalising symptoms using latent class growth analysis, a method that allows the estimation of different person-centred trajectories of symptom development. Subsequently, the studies investigate the socio-demographic correlates (including gender, socio-economic status, ethnicity and age) of individuals with different trajectories. This is followed by examining the impact of these different types of trajectories on another key domain of child functioning: educational attainment.

The fifth study focuses on the co-development of symptoms in the internalising and externalising domains in late childhood and early adolescence, trying to uncover patterns in their association and development over time in these two developmental periods.

The last study, based on the results of the previous studies, aims to investigate the underlying structure of child psychopathology. Using hierarchical bi-factor analysis, the study explores the possibility of a general propensity for psychopathology that not only is a better predictor of future psychopathology but can lead to better models of specific disorders as well.

The thesis contributes to increasing the understanding of complexities in symptom development and proposes a simpler structural model underlying symptoms. The analytic approach used highlights the value of modern data analytic techniques in increasing our understanding of the development of psychopathology.

Table of Contents

List of Figures

List of Tables

List of abbreviations

Below is presented a list of commonly used abbreviations in this thesis

BCAMHS	British child and adolescent mental health survey
CFA	Confirmatory factor analysis
CFI	Comparative fit index
CTT	Classical test theory
DIF	Differential item functioning
DTF	Differential test functioning
ES	Effect size
FSM	Free school meal eligibility
IRT	Item response theory
L-A-LOR	Liu-Agresti log odds ratio
LCGA	Latent class growth analysis
LGCA	Latent growth curve analysis
LRT	Likelihood ratio test
M&MS	Me and My School Questionnaire
MLM	Multi-level modelling
MLR	Multinomial logistic regression
RMSEA	Root mean square error of approximation
SDQ	Strengths and Difficulties Questionnaire
SEN	Special educational needs
TLI	Tucker Lewis index
WLSMV	Weighted least squares adjusted for means and variances

Publications and conference presentations associated with this thesis

Journal articles

Patalay, P., Fonagy, P., Deighton, J., Belsky, J., Vostanis, P., & Wolpert, M.(in press). A general psychopathology factor in early adolescence. *The British Journal of Psychiatry.*

Patalay, P., Deighton, J., Fonagy, P., & Wolpert, M. (2015). The Relationship between Internalising Symptom Development and Academic Attainment in Early Adolescence. *PLoS ONE, 10*(1), e0116821. doi: 10.1371/journal.pone.0116821

Patalay, P., Deighton, J., Fonagy, P., Vostanis, P. & Wolpert, M. *(2014).* Clinical validity of the M&MS questionnaire: a child self-report mental health measure. *Child and adolescent psychiatry and mental health.* doi: 10.1186/1753-2000-8-17

Patalay, P., Deighton, J., Fonagy, P., & Wolpert, M.*(advance online print).* Equivalence of paper and computer survey formats of a child self-report mental health measure. *European Journal of Psychological Assessment.* doi:10.1027/1015-5759/a000206

Conference presentations

'Testing the equivalence of paper and computer survey formats of a child self-report mental health measure using DIF analysis'. Paper presentation. Developmental methodology, Society for Research in Child Development. San Diego, USA. September, 2014.

'Same question, different methods, consistent answers? Examining the impact of externalizing symptom development on attainment'. Paper presentation. Developmental Methodology, Society for Research in Child Development. San Diego, USA. September, 2014.

'Re-examining the structure of psychopathology: a general psychopathology factor in early adolescence?'. Paper presentation. Child and Adolescent Mental Health Conference, Northampton, UK. July, 2014.

'Trajectories of internalising symptoms and links with academic achievement in primary and secondary schools'. Paper presentation16th European Conference on Developmental Psychology, Lausanne, Switzerland. September, 2013.

'Trajectories of emotional symptoms in primary school pupils in England: patterns and predictors'. Paper presentation.14th International Congress of the International Federation of Psychiatric Epidemiology, Leipzig, Germany. June, 2013.

Declaration of the candidate's role in each of the studies

Guidance and support were provided by both Prof. Peter Fonagy and Dr. Jessica Deighton throughout the entire thesis. Dr. Miranda Wolpert also provided input as PI of the wider study from which most of the data are drawn.

Most of the data utilised in the thesis, including all the data in Part 2, are from a wider study of mental health in schools in England – the 'Me and My School' study (see Wolpert et al., 2011 for the detailed report of the wider study). The study was led by Dr. Miranda Wolpert and the study team included Prof. Peter Fonagy and Dr. Jessica Deighton. The candidate's involvement commenced in the final phase of data collection of this wider study, and carried on to assist with analysis and interpretation of findings and preparation of the report.

Part 1. Background

The systematic review update reported in this chapter, used the protocols previously developed by Jessica Deighton and colleagues for the original review (Wolpert et al., 2009). All other work was the candidate's own.

Study 1. Clinical validation of the M&MS questionnaire

All work presented in this chapter was carried out by the candidate.

Study 2. Psychometric equivalence of the computer and paper survey formats of the M&MS questionnaire

The candidate conceptualised the study, gained ethics approvals for data collection and recruited participants. Data collection was carried out with assistance from other members of the

research team. Data for the computer survey sample were taken from the Me and My School study sample.

Part 2. Background.

All work is the candidate's own.

Study 3. Short-term externalising symptom trajectories: correlates and differential impact on academic attainment

All work is the candidate's own.

Study 4. Developmental trajectories of internalising symptoms: patterns, correlates and links with academic attainment

All work is the candidate's own.

Study 5: Co-development of internalising and externalising symptoms: a brief investigation

All work is the candidate's own.

Study 6: A general psychopathology factor in early adolescence

All work is the candidate's own.

General discussion

All work is the candidate's own.

Acknowledgements

Foremost, I thank my supervisors- Peter Fonagy and Jess Deighton- for your patience, guidance and nurturance. Your witty remarks, kind criticism and the challenging discussions we had helped shape this thesis and will always be part of my fond recollections of thesis writing days. There are many useful insights that I gained from both of you, some intended and some maybe not, that I hope will guide me through life, both research career and otherwise. I would like to extend this thanks to Dr. Miranda Wolpert for the intermittent yet always valuable insights and for making it possible for me to carry out this work.

I am extremely grateful to all the schools, parents and young people whose participation made this investigation possible and to the TaMHS wider research group for letting me join and be a part of the team (specifically to Profs. Panos Vostanis and Jay Belsky for their input on some studies).

I would like to thank Natasha and Amelia- collecting the TaMHS data would have been a much harder experience without your company, Emily and Dion for assisting me in collecting data for my studies and Eren and Andy for guidance with trajectory-based analysis and multi-level modelling respectively. I will never forget all my colleagues at the EBPU who make research always seem fun and am especially grateful to those I had the privilege of sharing an office with over the years for being patient enough to humour me during the rough patches and whom I can now gratefully count as good friends. I thoroughly enjoyed the entire process, which is not a minor achievement, and much credit goes to all of you.

I am extremely grateful to my parents Vani and Pradeep, for their support from the very beginning and my sister Anvitha for being a nuisance bundle of fun as often as possible. With all my heart I thank Martin for the many long conversations, being there through the ups and downs and showing me how it's done and maybe more importantly how not to do it.

This thesis would have not come about without all of your varied inputs and I hope I get the opportunity to share some of the findings with you.

Statement of funding

The wider study from which most data in the current thesis are drawn was funded by a grant from the Department of Children, Families and Schools (now the Department for Education) in England. The PhD was funded by a bursary awarded by the Evidence Based Practice Unit, University College London and the Anna Freud Centre.

Thesis Organisation

Two main areas can be distinguished in the study of developmental epidemiology- first, measurement and assessment of mental health and second, patterns of symptom occurrence and their predictors (Verhulst & Koot, 1992). The main themes of this thesis, as implicated by the title, revolve around the development and underlying structure of internalising and externalising symptoms in late childhood and early adolescence. The studies that fall within these themes are preceded by studies investigating the measurement properties of the key outcome measure used in the subsequent studies. Hence, the six studies in this thesis are presented in two parts with studies within each part (structure outlined in box below).

Part 1 pertains to the self-report measurement of mental health symptoms in young people and includes two studies whose aim is to further validate the self-report measure that is subsequently used in the studies that follow in Part 2.

Part 2 includes four studies which directly aim to contribute to the two main aims of the thesis that have been outlined above. Studies 3 and 4 investigate the complexity in development of externalising and internalising symptoms respectively. Study 5 examines the associations between externalising and internalising domains over time and the last study (Study 6) explores the possibility of a general psychopathology dimension in young people.

Part 1: Self-report measurement of mental health in school aged children: further validation of the Me and My School (M&MS) questionnaire

Background

Study 1: Clinical validation of the M&MS questionnaire

Study 2: Psychometric equivalence of the computer and paper survey formats of the M&MS questionnaire

Part 2: Externalising and internalising symptoms: developmental trajectories, co-development and underlying structure

Background

Study 3: Short-term externalising symptom trajectories: correlates and differential impact on academic attainment

Study 4: Developmental trajectories of internalising symptoms: patterns, correlates and links with academic attainment

Study 5: Co-development of internalising and externalising symptoms: a brief investigation

Study 6: A general psychopathology factor in early adolescence

General discussion

Part 1: Self-report measurement of mental health in school aged children: further validation of the Me and My School (M&MS) questionnaire

Background

The estimated prevalence of mental health difficulties experienced by children and adolescents ranges from 10-20% (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; Green, McGinnity, Meltzer, Ford, & Goodman, 2005). The most common mental health problems in young people are conduct problems and emotional problems followed by hyperkinetic disorder and less common disorders such as eating disorders and autism (Green et al., 2005). A recent school based survey of mental health in England indicates that, on average, in primary schools in England up to 9% of children have emotional problems and 9% behavioural problems and in secondary schools up to 6% of children have emotional problems and up to 8% have behavioural problems when measured via self-report (Wolpert et al., 2011).

Schools are the community place outside the family where children spend most of their waking hours (Jacquelynne S. Eccles & Robert W. Roeser, 2011). Furthermore, schools are an ideal setting for identifying young people with mental health problems (Levitt, Saka, Romanelli, & Hoagwood, 2007) and the structure of the educational system provides an easy sampling frame and allows access to large numbers of young people (Oppenheim, 1992).

Mental health difficulties not only effect the well-being of young people but also affect their school work (Masten et al., 2005) and social relations (Eccles & Roeser, 2011). Furthermore, children with such problems can cause disruption in classrooms and affect others in school, for example a child with behavioural problems might bully other children and cause distress and emotional problems in these children (Olweus, 1993). Studies have shown that a majority of school aged children with mental health problems are not identified and treated by appropriate services (Costello, Mustillo, et al., 2003; Ford, Hamilton, Goodman, & Meltzer, 2005; Kataoka, Zhang, & Wells, 2002). It is desirable that children with mental health difficulties

are identified early and given the right intervention/treatment (Levitt et al., 2007) as children with mental health problems have a higher risk of growing up with anti-social behaviours and other health problems (Broidy et al., 2003).

Schools can use the information from child mental health measures to screen and identify children who might be at risk or experience difficulties and require intervention (Levitt al, 2007). Measuring mental health in children can also be used to assess the impact of an intervention implemented within or outside the school. Measuring mental health in schools is also useful at a more aggregated level for research purposes to help assess prevalence of problems, effectiveness of treatments and interventions which will help inform policy and the health services (e.g., Wolpert et al., 2011).

Different reporters of mental health in school aged children

Mental health in school children can be measured from the perspective of various reporters which include the young person themselves, a parent, teacher, carer or a peer. It is recognised that individuals have a unique perspective of their own emotional state and quality of life, which depends among other things on their past experience, present lifestyle and expectations and aspirations for the future (Eiser & Morse, 2001). The different insights and perspectives that are obtained from different reporters depend on factors such as the relationship between the reporter and the young person and the context and duration of their contact with the child (Achenbach, McConaughy, & Howell, 1987; Verhulst & van der Ende, 2006). Research indicates that children and parents often have different perspectives on health (Eiser & Kopel, 1997), and their reports on mental health and quality of life often differ substantially (Cremeens, Eiser, & Blades, 2006). For instance, parents' views might be influenced, among other things, by their own mental health and concerns about their child's illness (Eiser & Morse, 2001). Yet,

parent reports are most often used to report on the mental health of children and adolescents (Bullinger & Ravens-Sieberer, 1995). Teachers' report of child mental health are also used, especially in the school context as the teacher is expected to be aware of the child's behaviour and emotional state at school. The feasibility of using teacher reports for a large number of pupils is limited and depending on class size and interaction with individual pupils' teachers may or may not be aware of children's difficulties, especially emotional difficulties (Kolko & Kazdin, 1993).

Accuracy of different reporters has also been shown to be different depending on type of psychopathology that is being measured (Verhulst & Van der Ende, 2009). Evidence suggests that parents and teachers are better at recognising and reporting on behavioural difficulties of school children but not equally accurate at identifying children with emotional difficulties, a finding that is attributed to the observable nature of externalising problems (De Los Reyes & Kazdin, 2005; Kolko & Kazdin, 1993). Agreement between parents and adolescence on symptoms of internalising problems, especially depression, is very low (Cantwell, Lewinsohn, Rohde, & Seeley, 1997). Children are generally considered better reporters of internalising problems, since internal states might be harder to perceive for other individuals, such as parents and teachers (Cantwell et al., 1997; Verhulst & Van der Ende, 2009) Hence, the child's perspective is key to identifying children with difficulties, especially emotional difficulties and some recommend the use of child self-report when only data from a single reporter is possible or feasible (Cantwell et al., 1997).

However, there are potential limitations with children self-reporting that researchers have identified. Some of these limitations pertain to children completing questionnaires in general and some are more specific to measuring concepts such as mental health, well-being and adjustment.

The limitations of children responding to questionnaires in general have been summarised by Schmidt, Garratt, and Fitzpatrick (2001) as 1) position bias (tendency to select first response), 2) acquiescent response bias (tendency to agree with all statements), 3) limited understanding of negative and complex items, 4) problems perceiving time periods and 5) limited comprehension of complex sentences and concepts (Schmidt et al., 2001). With regards to children specifically answering mental health questionnaires Wolpert et al. (2008) have outlined the following limitations: 1) children might be more likely to give socially desirable answers about their own mental health and behaviour (when compared to other reporters) 2) children with problems might have low self-awareness regarding these problems, and 3) children might be less consistent in their self-perception regarding mental health difficulties and might not respond based on stable adjustment but rather the 'here and now' (Wolpert et al., 2008).

In terms of these limitations to children answering questionnaires, some of these effects can be limited by asking simple questions which are easy to understand and do not have complex sentence structure (Bullinger & Ravens-Sieberer, 1995). Acquiescent response bias is also seen in adult reporters and is not limited to children self-reporting. Carefully planning items and having positive and negative items can help limit position bias and acquiescent response bias (Ray, 1983). In regards to the social desirability bias, this is also true of adult reporters self-reporting behaviours and constructs such as mental health and hence is not specific to self-reporting by children (Grimm, 2010). Anonymity and privacy might go some ways in limiting desirability biases but it is a universal limitation of measures (Furnham, 1986), and as it is not limited to children self-reporting it should not be considered a deterrent to using child self-report measures. In regards to the limitation 'low self-awareness', it is true that some children might not realise the extent or nature of their problems but as mentioned earlier other reporters such as parents and

teachers are less likely to be aware of children's emotional problems and hence it is important to get the child's perspective to prevent children with emotional problems being missed out. In regards to the 'here and now' limitation, this is more likely in younger children and relates to the general problem with younger children perceiving time periods (Schmidt et al., 2001). To some extent this might be combated by making clear in questionnaires the time period of reference.

An assessment of the potential limitations of child self-report indicates that the limitations are not hugely different from the limitations of adults self-reporting. Some of these limitations can be reduced by good questionnaire design and being aware of the development stage and capacity of the children who are expected to use an instrument. The other limitations that cannot be controlled for through good measure design pertain more to the young developmental stage of children. This should be borne in mind while designing instruments and evaluating the data obtained from children. Bullinger & Ravens-Sieberer (1995) make the point that children are able to self-report on their health or quality of life provided the instrument used is appropriate to the child's ability (Bullinger & Ravens-Sieberer, 1995). The age at which children can reliably self-report might differ based on development and the construct of interest, for instance children as young as 4-5 years can report on pain, whereas for feelings and behaviours (internalising and externalising) this might be 5-7 years and possibly even older for a concept such as quality of life (Arseneault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005; Schmidt et al., 2001; Sharp, Goodyer, & Croudace, 2006). However, there is increasing evidence that even children with (severe) mental health problems understand and have insight on their difficulties and can provide information that is unique and informative (Schmidt, Garratt, & Fitzpatrick, 2000).

Additionally, there is increasing emphasis on children's views and their perspectives being taken into account. This was initially recognised by the UN Convention on the Rights of

the Child ("UN Convention on the Rights of the Child," 1989) and can be seen in national policies and legislation such as Every Child Matters (Department for Education, 2003) and Children Act (2004), advice from the department of health regarding patient reported outcomes (Department of Health, 2010). Apart from being better at reporting on internalising difficulties and having a good understanding of their own problems (Kolko & Kazdin, 1993; Schmidt et al., 2000), child self-report measures have practical advantages. They are more efficient for population based surveys as young people can themselves report on their health, even at several time points, and this results in less administrative and time burden for researchers and other reporters such as teachers, parents and clinicians (Wolpert et al., 2008). Moreover young people self-reporting on their own mental health allows for longitudinal measurement over time that is consistent as for e.g., in school settings pupils' teachers change at least yearly, if not more often, which makes for inconsistent reporting.

Upon assessment the limitations of children self-reporting are outweighed by the advantages and children can therefore, in many cases, be considered key and able reporters of their mental health. The following section reviews existing self-report measures of general mental health for use in community settings such as schools.

Review of existing measures

The aim of this section is to identify already existing self-report measures of mental health for children and adolescents. This was carried out in two stages: 1. A review of existing reviews of mental health measures for children and young people was undertaken and 2. An existing systematic review of child and adolescent mental health and well-being measures, Wolpert et al. (2008) was systematically updated.

Following the reviews identifying existing measures, inclusion and exclusion criteria were applied to the identified measures to filter down to measures that were self-report, designed for school aged children, measured broad mental health constructs and have been used in normative populations.

Review of reviews

Existing reviews of mental health and well-being measures for children and young people were scanned to find child self –report measures of general mental health and well-being (as opposed to measures of specific disorders/problems). The following reviews were identified based on searches of online databases, books and reports.

1. Review and recommendation for national policy for England for the use of mental health outcome measures with children and young people (Wolpert et al., 2008)
2. Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments (Solans et al., 2008)
3. Assessment Scales in Child and Adolescent Psychiatry (Verhulst & van der Ende, 2006)

4. Mental Health Outcomes Compendium (National Institute for Mental Health in England, 2008)

Based on the reviews, focussing on the results for generic mental health self-report measures, a list of measures was derived. Due to the nature of the reviews and the criteria used to select measures the results of the various reviews were largely similar. Of the existing reviews, Wolpert et al. (2008), was systematic and fully covered the aims of the existing review of measures. Hence, instead of starting a systematic review from scratch, the Wolpert et al. (2008) review was systematically updated using the same protocol of the existing review. This was done to establish the existence/lack of newer measures that met the criteria for a validated self-report measure of general mental health. The identified measures from the above reviews and the systematic review update are listed in the sections below, where subsequently further inclusion and exclusion criteria are applied.

Systematic review update

An existing review of mental health measures for young people Wolpert et al. (2008) was systematically updated (2008-2012) using methods, databases and key words that were identical to the existing systematic review to ensure continuity and good method, (detailed systematic review in Appendix A). Stage 1 involved searching four databases MEDLINE, EMBASE, ERIC and PsychInfo by keywords related to measures, mental health and well-being and children and adolescents. Searches that resulted in more than 100 hits were subject to basic filtering and excluded if title was not related to child and adolescent mental health outcome measures. In Stage 2 the remaining papers were filtered by applying the following exclusion criteria to the title and abstract: No child or adolescent mental health outcome measure mentioned, the measure was narrow and focussed on one disorder, the paper referred to a measure not used with children and the paper referred to a diagnosis. This resulted in 20 measures being identified.

In Stage 3 detailed inclusion and exclusion criteria were applied to the measures. The main inclusion criteria included were that the measure had to be available in English, measured broad mental health symptoms and not a specific diagnoses, took less than 30 minutes to complete, could be used with a fairly broad age range, had been validated with young people. The main exclusion criteria included: completed only by a professional (clinician), not used with children, open ended responses that have to manually coded, and measures that the existing review had already identifies were also filtered out to prevent repetition. Four measures were identified after Stage 3: the Brief Problem Checklist (BPC), the Diagnostic Infant and Pre-school Assessment (DIPA), QBH-16 (questionnaire based on HEADSS approach) and the Brief Child and Family Phone Interview (BCFPI) measures which did not have enough information in the papers to make a decision regarding inclusion or exclusion remained included at this stage.

In Stage 4 these measures were explored in more detail and measures identified as meeting the exclusion criteria were filtered out (details in Appendix A) and one measure remained: Brief Problem Checklist (BPC, Chorpita et al., 2010), which was further explored. The BPC is a shortened version of the self-report measure in the ASEBA and is has recently been developed and validated in clinical settings for use as a clinical outcome measure. It was excluded as the original long measure is already included in the shortlisted measures below. Additionally, having been solely validated in clinic samples so far it does not meet the criteria for use in community-based samples.

Shortlisted Measures

As the 2008-2012 systematic review update did not result in any additional new measures being added to the list of measures identified by Wolpert et al. (2008), based on that review and the other existing reviews of measures, 10 self-report measures are included below. These 10 self-report measures identified have been used widely to measure well-being, mental health or quality of life and have shown to be reasonably reliable and valid.

1. Youth self-report (YSR) from ASEBA
2. Child Health Questionnaire (CHQ) Child Form
3. Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) Self Rated
4. Kidscreen
5. Strengths and difficulties questionnaire (SDQ) Self Report
6. Youth Outcome Questionnaire (YOQ) Self Report
7. Becks Youth Inventories (BYI)
8. Youth Rating Scale from Behavioural and Emotional Rating Scale (BERS-2)

9. Pediatric Symptom Checklist- Youth report (PCS-Y)
10. Self Report of Personality (SRP) from Behaviour Assessment Scales for Children (BASC)

Based on the aims of the current review: to identify measures that are self-report, designed for school aged children, measured broad mental health constructs and have been used in normative populations exclusion criteria were applied to filter out measures that did not meet the above criteria. Considering the focus on a school based setting, measures that did not meet the criteria of having been used or validated in a normative population (not just a clinical population) such as the HoNOSCA, YOQ and PSC were excluded. The remaining seven measures were then assessed, based on both psychometric properties and practical considerations, within the measure assessment framework that is outlined in the next section.

Measure Assessment Framework

For the purpose of assessing/evaluating existing measures a framework has been outlined in this section within which the properties of existing measures will be assessed. Traditionally most instrument assessment has focused on mainly concepts from the traditional psychometric framework (such as reliability, validity and generalisability) but it is increasingly recognised that measure assessment criteria must include consideration of other aspects of instruments such as item construction, biases, settings, practicality and feasibility (Scientific Advisory Committee of the Medical Outcomes Trust, 2002).

The framework for assessment of measures established and described by the Scientific Advisory Committee (SAC, 2002) for assessing health status and quality of life instruments was chosen as it integrates traditional psychometric criteria, influences from modern item response

theory and also criteria relating to measure design and practicality of usage. This was deemed important to include because, as previously outlined - item construction and content play a major role in determining the suitability and functionality of child self-report measures. The framework was developed by the Scientific Advisory Committee as part of their role in reviewing instruments and assessing their suitability for widespread use (SAC, 2002). Some aspects of this framework have been further detailed by Terwee et al. (2007). The criteria described below have input from the main framework (SAC, 2002) and where information was sparse, the details and guidance from the follow up paper (Terwee et al., 2007) have been used. The authors of the framework suggest eight main attributes that allow an overall assessment of measures. The importance of the eight attributes varies depending on what aspect of instruments is of interest, for example criteria of alternate forms and language adaptations only become relevant when these aspects are involved, whereas reliability and validity are key concepts that are always important to take into consideration. These eight attributes and their application to child self-report mental health measures are described below.

1. Conceptual and measurement model

The conceptual model of a measure has been defined by SAC (2002) as the rationale for and description of the concepts and the populations that a measure is intended to assess and the relationship between these concepts. Assessing this attribute includes considering the conceptual and measurement basis for item inclusion and combinations, target group involvement in questionnaire design, information on dimensionality and distinction of scales, evidence for scale variability and rationale for deriving scale scores. Within an Item Response Theory framework, Rust and Golombok (2009) discuss how differential item functioning (if items operate differently across groups) and item difficulty (the extent to which the item discriminates individuals on the

construct being measures) analysis can be used to assess scale variability across groups and forms a measurement basis for item inclusion and exclusion, to ensure items in scales measure the same thing in different sub-groups of the target populations (Walker, 2011).

As discussed previously, when designing child measures it is vital to bear concepts being measured and responding population (children) in mind as children might have different conceptualisations of mental health than adults and different aspects of this might be salient to them. The level of complexity of sentences and construct that children in the target ages can comprehend must also be taken into account. Involving children either as collaborators or as commentators on the items, item selection and concepts covered would be one way of ensuring instruments are suitable for child self-report.

2. Reliability

Reliability can be explained as the degree to which a measure is free from random error and this can be assessed by looking at the internal consistency and reproducibility of measures. Anastasi & Urbina (1997) consider that reliability underlies the computation of error of measurement of scores obtained from measures and thus knowing reliability indicates the extent to which individual differences represent 'true differences' and not just measurement error. Different aspects of reliability can be assessed such as internal consistency and reproducibility.

Internal Consistency. The extent to which items in a (sub) scale are inter-correlated, thus measuring the same construct. Cronbach's alpha (Cronbach, 1951) is most commonly used as the measure of internal consistency of a scale. Anastasi & Urbina (1997) point out that alpha is dependent on content sampling and the nature (heterogeneity/homogeneity) of the domain being sampled. In measuring mental health domains in children this would be an indication of how

homogenous items in a scale are as to what domain they are measuring. Hence, low internal consistency (Cronbach's alpha < 0.70) could be an indication that the items in a scale are not measuring a similar domain and hence aggregating them might not be justifiable (Anastasi & Urbina, 1997). In regards to existing child mental health, measures Verhulst & van der Ende (2006) discuss that even though internal consistency for whole scales (total scores) is usually acceptable, some subscales can have low reliability and that this can be problematic (Verlhust & van der Ende, 2006) .

Reproducibility. Also known as test-retest reliability this criteria looks at the extent to which repeated measurement in stable persons (test-retest) provide similar answers. Terwee et al., present criteria for justifying the assumptions of a test-retest analysis such as correlations must be presented along with a description of data collection and the time period between the tests and a justification of why the population might be considered stable across the time period. In regards to child mental health measures, as previously mentioned, it is believed that children tend to answer questions about their mental health based on the present moment rather than on the general state of things. This characteristic of younger children answering questionnaires might lead to lower reproducibility as there might be less stability in scores, even over short time periods. Verlhust & Van der Ende (2006) have suggested that the time intervals used for the test-retest reliability for child mental health measures should be short enough to expect that the subject's behaviour did not change in that time period and suggest periods of one to two weeks between testing periods.

3. Validity

Validity of a measure can be described as a measure of the degree to which the instrument measures what it purports to measure. SAC (2002) include three sub-types of validity testing as

key considerations with health status and QoL related scales. This can be assessed by looking at various types of validity including content, criterion and construct validity.

Content Validity. SAC define content validity as the extent to which the domain of interest is comprehensively sampled by the items in the questionnaire. Terwee et al. (2007) & SAC (2002) lay out the following aspects of measure development by which this can be assessed: 1) clear description of the aim, target population, the concepts being measured and the item selection procedure (including item reduction techniques) must be provided. 2) The target population must be defined and they should be involved in item selection. 3) Interpretability of items should be clear which means that items should be short and clear, not consist of two questions at the same time (double-barrelled items) and the time period to which the items refer to must be clearly stated. As discussed in the background, these considerations are especially important when constructing child self-report measures as children have more trouble understanding complex, double barrelled items with complex response options.

Criterion Validity. Criterion validity is often described as the extent to which scores on a measure relate to a 'gold standard'. To evaluate this there must be sufficient justification that the gold standard is actually 'gold' and then the extent to which the measure correlates with this standard is assessed. In mental health measurement the lack of a gold standard stems from the need to incorporate information from multiple sources and the difficultly to assess whether a behaviour constitutes as abnormal due to the subjectivity of the construct being measured (De Los Reyes & Kazdin, 2005; Verhulst & van der Ende, 2006).

Construct Validity. The extent to which scores on a questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that

are being measured. As Campbell (1960) noted, this must not only include high correlation with variables with which it should theoretically correlate but also low correlation with variables from which it should differ i.e. convergent and discriminant validation (Campbell, 1960). For instance, for two measures of mental health in children one would expect the emotional problems scales to correlate highly whereas the behavioural problems scale should have a low correlation with the scale measuring emotional problems. Terwee et al. (2007) also include expected differences in scores between 'known groups' as a part of construct validity. In regards to mental health measures for children differences between known groups can include differences between clinical and normative samples, children with special educational needs and children with record of deviant behaviours.

4. Responsiveness

Responsiveness is also known as sensitivity of a measure and it refers to a measures ability to detect change over time (SAC, 2002). Sensitivity is an important attribute of health status measures as they are increasingly used to assess the impact of treatments or interventions (Verhulst & Van der Ende, 2009). Considering that it is linked to a measures ability to measure reliable change, it is inherently linked to the validity and reliability of a measure, but a measure can be highly reliable but not sensitive to change and vice versa (Guyatt, Walter, & Norman, 1987). Guyatt et al. (1987) proposed calculating' minimal clinically important difference' (MIC), however clinically important change in mental health can be difficult to empirically define (Katz, Larson, Phillips, Fossel, & Liang, 1992). Terwee et al. (2007) suggest that sensitivity in a clinical context, can be assessed by calculating the differences between groups that are known to have and have not changed or comparing scores of relevant sub-groups that are expected to differ (Terwee et al., 2007) and Katz et al. (1992) use standardised response mean to calculate change

(Katz et al., 1992). Sensitivity of child mental health measures is not commonly assessed and reported, particularly when first validated. However, this is an especially relevant criterion in child mental health measures used in clinical practices as it would be desirable that a measure detects if a child is improving or not and this in turn will help assess efficacy of treatments and interventions (Verhulst & Van der Ende, 2009).

5. Interpretability

Interpretability has been explained by SAC (2002) to be the degree to which one can assign easily understood meaning to quantitative scores, and this is the feature of measures that allow qualitative interpretation of quantitative scores on different scales (Terwee et al. , 2007). They propose criteria for interpretability that include 'benchmarks' to facilitate interpretation of scores, rationale and comparisons of population and information regarding how the data from an instrument should be calculated and reported (SAC,2002). In mental health measures used in population settings this usually takes the form of sub-scales of items measuring different constructs, cut-offs that suggest clinical levels of problems and different norms for different groups. Anastasi & Urbina (1997) argue that for norms to be used they must be calibrated for the same population as these are usually sample dependent. They explain that within group norms (same age, gender or any other criteria used) can be helpful to plot scores and compare to the population of reference (Anastasi & Urbina, 1997). For child mental health this could take the shape of being able to compare levels of difficulty of individuals or samples in the context of a similar population and this is important as the nature of mental health problems and the way they manifest differ across different periods of development (Verhulst & Van der Ende, 2009). Additionally, the extent to which scales translate to existence of disorder as diagnosed by

professionals can help determine the interpretability of scores from mental health scales or sub-scales.

6. Respondent and administrative burden

The time, effort, and other demands placed on those to whom the instrument is administered (respondent burden) or on those who administer the instrument (administrative burden).

Respondent burden. Respondent burden includes consideration of time and reading and comprehension level required and strain on the respondent. For child measures it is important to consider that children have slower reading and comprehension times and questions that seem quite straightforward to adults could place higher amounts of strain on children. It is important to make both items and response options as simple and easy to understand as possible to minimise respondent burden. Another way to reduce burden would be to minimise the number of items while ensuring that there are enough items to capture concepts being measured.

Administrative Burden. Considerations of administrative burden include requirement of resources, time required to administer questionnaire and training and expertise required to interpret the outcomes. Especially For epidemiological or large scale measurements of mental health Verhulst and Van der Ende (2009) note that ‘’assessment scales need to be accurate, practical and economical’’.

7. Alternative forms

Alternative forms can include different formats of the same measure (computer, paper, interview) and also different reporters of the same measure (self-report, parent report etc.). This can be assessed by available evidence on reliability, validity, administrative and respondent

burden for alternative forms and also information on the comparability of the alternative forms of measures. For child mental health measurement in schools the availability of alternate forms is desirable as for large population based studies in school settings a computer based measure reduces the burden of data collection (e.g., Wolpert et al., 2011) but in other circumstances such as for use by counsellors and school staff a paper based version might be more suitable.

8. Cultural and language adaptations (translations)

In addition to possibilities of cross-national/cultural studies and comparisons facilitated by measures being validated in multiple languages, in a multi-cultural society, such as Britain today, availability of a measure in different languages and sensitivity to different cultures is being seen as a key aspect to make a measure more widely available and acceptable (NIMH, 2008). Merely translating measures is not sufficient and assessing translations of measures involves two main steps, 1) assessment of conceptual and linguistic equivalence and 2) evaluation of measurement properties of the translations (SAC, 2002).

Review of existing measures within the measure assessment framework

The measures identified in the previous section based on the reviews were then assessed against the criteria in the measure assessment framework outlined in the above section. Table 0.1 contains the mapping of the identified measures with each of the criteria and a summary of how each criterion is met based on existing validation papers and manuals of each of the measures.

Table 0.1. The seven measures identified by the review set against the criteria described in the measure assessment framework

Assessment criteria	**Youth self-report (YSR)**	**CHQ Child Form**	**Kidscreen**	**SDQ Self Report**	**Becks Youth Inventories (BYI)**	**Youth Rating Scale (BERS-2)**	**Self Report of Personality (BASC)**
Measurement model	Covers eight domains of mental health. Designed and validated with children ages 11 years and above	Covers 10 domains across physical and mental health	Health related quality of life measure covering domains such as parent relations, peers, school environment	Mental health measure with 5 subscales. Designed and validated with children ages 11 years and above	Includes 5 inventories of anxiety, depression, anger, disruptive and self-concept. Designed and validated with children aged 7 years and above	Covers 6 domains. (e.g., Interpersonal Strength, Family Involvement) Designed and validated for children aged 5-18 years	Covers 14 domains such as hyperactivity, attention problems etc. Designed and validated for children aged 11 years and above.
Internal consistency	0.55-0.95	0.62-0.94	0.77-0.89	0.41-0.81	0.86-0.92	0.79-0.95	0.80-0.82
Reproducibility	0.68-0.91 (8days)		0.7 (2 weeks)	0.21-0.62 (4-6 months)	0.63-0.89 (1 week median)	0.84-0.91 (2 weeks)	0.64-0.86 (1 month)
Content validity	Item selection reported	Item selection reported	Item selection reported	Item selection not reported. Contains complex, double barrelled items	Item selection reported	Item selection reported	Item selection reported
Construct validity	CBCL and TRF	Parent 25 and Parent 50	KINDL scales 0.16-0.68; Peds QL 0.44-0.61	SDQ parent and teacher version	CDI 0.26-0.72; RCMAS scales 0.13 - 0.7; PHCSCS scales 0.06 - 0.67; CASS:S 0.27 - 0.73	YSR 0.03-0.81; SSRS 0.32-0.73	MMPI (0.78-0.89)

Interpretability	Discriminates between referred and non-referred samples Clinically relevant cut-off's	Discriminates between community and clinical samples. No available norms yet.	Discriminates between healthy and mentally or physically ill children	Discriminates between clinical and normative samples. Clinically relevant cut-off's	Discriminates between clinical groups and controls; SEN groups and controls Clinically relevant cut-off's	Discriminates between normative sample and sample with emotional and behavioural problems	Discriminates between different clinical profiles.
Respondent burden	105 items	87 items	52, 27 or 10 items	25 items	100 items	57 items	139-185 items
Admin burden	Costs	Costs, free for research	Free	Paper copies are free; costs for online use	Costs	Costs	Costs
Alternate survey formats and language adaptations	Paper and computer Approx. 50 languages	Paper and computer. Approx. 25 languages	Paper and computer Approx. 25 languages	Paper and computer Approx. 50 languages	Paper and computer English	Only paper English and Spanish	Paper and computer English and Spanish

References: YSR (Achenbach & Rescorla, 2001); CHQ (Landgraf, Abetz, & Ware, 1999); Kidscreen (Ravens-Sieberer, Auquier, et al., 2007; Ravens-Sieberer et al., 2010; Ravens-Sieberer, Gosch, et al., 2007); SDQ (Goodman, 2001; Goodman, Meltzer, & Bailey, 1998); BYI (Beck, Beck, & Jolly, 2001); BERS-2 (Epstein, 1999, 2004; Epstein, Harniss, Ryser, & Pearson, 1999); BASC (Flanagan, Alfonso, Primavera, Poval, & Higgins, 1996; Flanagan, 1995; Reynolds & Kamphaus, 1992)

In terms of measurement model all seven measures cover a wide range of domains between them. Some measures such as the SDQ and the SRP cover domains that directly measure mental health outcomes such as emotional and behavioural problems whereas some of the measures such as the Kidscreen and CHQ cover all aspects of health including physical health and the BERS-2 covers domains that are more related to functioning such as family relationship and school functioning. However, the Kidscreen does not include any behavioural domains covering externalising symptoms, which considering it is a highly prevalent, salient problem in schools limits the utility of this measure for school based mental health screening.

As can be seen from the table, internal consistency of the measures varies greatly between the measures. The Kidscreen, BYI, BER-2, and BASC have acceptable alphas according to the widely used criteria of above 0.7 being good. On the other hand the SDQ and YSR have some scales with quite poor alphas (e.g., the SDQ conduct scale has an alpha as low as 0.45 in some studies (e.g., Muris, Meesters, & van den Berg, 2003).

Reproducibility of most of the measures measured at around 1 week ranged between 0.63 – 0.91 across the scales of the different measures which, considering the possibly higher volatility of children's responses that was discussed earlier, is reasonable in most instances, especially considering that 0.7 is recommended as a minimum standard of reliability (Terwee et al., 2007).

Content validity will be discussed within the three sub-heading that were laid out in the framework above that are taken from SAC (2002). The first point relates to a clear description of the aim, target population, concepts and procedure of item selection. All seven measures have clearly defined aims and concepts being measured and all but the SDQ have reported in detail the process of item selection. The second point within content validity

relates to the involvement of target populations in item selection; most of the measures have involved target population up to some extent in item selection, most commonly in the first stage of item generation (e.g., YSR, Kidscreen, BYI) and some have systematically tested items with target populations (e.g., Kidscreen). Again the SDQ stands out as being the only measure that has not reported target population involvement in item selection. The third point under content validity includes considerations of item clarity and interpretability. According to Terwee et al. (2007) this means 'items should be short and simple and not contain any difficult words or jargon terms. Moreover, items should not consist of two questions at the same time'. The former is less easy to assess as there are no criteria laid out for what constitutes short and simple items. However, in terms of double-barrelled items, for e.g., the SDQ has nine (out of 25) double barrelled items (e.g., I fight a lot. I can make other people do what I want). Most of the other measures in this review avoid complex double-barrelled items.

Construct validity of the measures was assessed by comparing the scale correlations of the measures to other measures and as outlined earlier, convergent validity of like scales and divergent validity of dissimilar scales were considered as indicating acceptable construct validity. All seven scales have shown to have satisfactory construct validity and no scale stands out as having particular limitations in regards to this kind of validity. As discussed in the previous section in mental health, there is no 'gold standard' that measures can be compared to so comparisons between measures and with other indicators are the only criteria against which construct validity can be assessed. It is important to note that many of the measures were assessed for construct validity by being compared with the parent or other reporters of the same measure, rather than other self-reported measures.

In terms of whether the measures discriminate between clinical and normative populations, all the measures have shown evidence of discriminating sufficiently between

populations that would be expected to differ. Most of the measures have been shown to discriminate between healthy and clinical or referred (e.g., YSR, SDQ and BERS-2) samples and some such as the SRP have also shown to discriminate between different clinical profiles.

Length of the questionnaire was looked at as an indication of respondent burden on children completing the questionnaires. In terms of length five out of the seven measures have more than 50 items and Kidscreen (27 & 10 item versions) and SDQ are the only measures that have relatively lesser number of items.

Administration burden was mainly assessed by whether a measure had to be paid for or was free to use. Out of the seven measures being looked at here, only two were free to use as paper versions (SDQ and Kidscreen) and only one, the Kidscreen was free to use in all formats. The CHQ is free only for research purposes. All the other measures (YSR, BYI, BERS-2 & BASC) cost to use.

In terms of alternate forms that can be used most measures have computer based and paper-pencil formats, except the BERS-2 which reports having only a paper version of the measure. However, none have validated and explored the equivalence of the alternative survey formats. Most of the measures, except the BYI, are also available in multiple languages with the YSR, CHQ, Kidscreen and SDQ being available in more than 25 languages each.

Based on the attributes of the measures outlined here, the following section will assess the suitability and limitations of these measures as child self-report, school based measures of mental health.

Limitations of the existing measures

The above measures are all widely used and validated measures of different aspects of child well-being and mental health. In the following discussion, based on the criteria discussed in the previous section and how all the measures fare within the assessment framework (discussed above and in Table 1 above), they are evaluated in terms of their suitability for a large population based study in primary and secondary school settings. Verlhust & Van der Ende (2006) point out that for large scale population based studies measures need to be 'practical, accurate and economical'. The limitations of each measure for this purpose are discussed below with the aim of supporting the creation and further validation of a school based measure of mental health.

The YSR, BYI, YRS and SRP all have more than 100 items which increases respondent burden, they cost to use which is an additional administration burden and have been used and validated only with children ages 11 years and above (except BYI which is from age 7) which limits it to only secondary school pupils. The copyright and cost issues also prevent creation of shorter or simpler versions of these measures which might have been suitable to use with primary school aged children as well.

The CHQ is a shorter measure compared to the four above (87 items) and free for research purposes but the main limitations (in terms of our purpose) is that it has only been used and validated with children aged 11 and above and the higher number of items increase respondent burden and make it unsuitable to adaptation with younger children. Also, as the name suggests, it measures more broad concepts of health alongside mental health.

The Kidscreen is a measure of quality of life that has been validated to use with children aged 7 and above. It has strong psychometric properties and is a well validated measure. The key limitation of this measure in terms of mental health in schools is that it does not have scales that measure behavioural problems. It has a few scales that measure emotion-

related difficulties but since behavioural problems are of key importance in a school setting the lack of a scale that measures behavioural difficulties makes it unsuitable for a school based study of broad mental health difficulties.

The SDQ is a measure developed in England and is widely used for mental health research and in clinical practice in the UK. It has been designed and validated for children aged 11 and above and it cannot be adapted to younger children as the items and item structures are too complex for a younger reading age. It also includes double-barrelled items (e.g., I can fight a lot. I can make other people do what I want) which is not considered good practice in measure design as a positive answer does not clarify which part of the question is being answered to, especially if the two sentences in the item may be understood slightly differently (Oppenheim, 1992). Another limitation of the SDQ is that the conduct problem scale has very low internal reliability and the whole measure has low reproducibility.

In summary, a review of existing measures of child mental health suggests that there is no free-to-use, brief child self –report measure of general mental health that can be used with children aged under 11 years in primary schools in community based settings, for instance a large school-based study, and as a screening tool. To fill this gap, the Me and My School (M&MS) was created for use in a national evaluation of mental health provision in primary and secondary schools in England. The measure was created to be an online, computer based survey and was for the greater part used in this format for a large study of mental health and provision in schools in England (Wolpert et al., 2011).

The Me and My School Questionnaire

The Me and My School questionnaire (M&MS, Deighton et al., 2013) is a self- report measure of young people's mental health that was designed to act as a screening measure in a school based setting. It is suitable to use and has been validated with children as young as eight years old, which makes it (as far as the authors are aware), the only free to use, validated brief self-report measure of general mental health for children as young as eight years. Initial validation and analysis of psychometric properties has revealed it to be a measure with good content validity, internal reliability, construct validity and minimal item-bias. The measure has 16 items and takes 5-10 minutes to complete. The items have simple sentence structure and response options for the items are on a 3-point Likert scale: Never, Sometimes and Always.

Deighton, Tymms, et al. (2013) demonstrated in the validation study in a large community sample that the measure has good content validity, as indicated by the item-selection procedure. Young people were involved in the development of the measure, which resulted in items that are simple and easy to understand, which makes it suitable to use by younger age groups. The scale structure clearly divides the measure into two related yet symptomatically diverse scales, emotional difficulties and behavioural difficulties, corresponding to the domains of internalising and externalising symptoms respectively. Construct validity was established using an existing widely used measure, the SDQ (Goodman et al., 1998), and the final scale has items that do not show differential functioning based on developmental stage, socio economic status and special educational needs (Deighton et al., 2013). Thus far the assessment of the measure has fulfilled criteria including measurement concept, internal reliability, content validity, and construct validity.

To establish the properties of this measure in accordance with the criteria outlined above further validation of the measure is required. Two additional studies will be presented here that further the validation process of this measure.

Study 1 aims to assess if the measure discriminates successfully between a normative community sample and a clinical sample. The results will give us more information on the discriminant validity and the interpretability of the measure.

Study 2 aims to validate the paper version of the measure. This will meet the criteria to assess alternative forms of the measure. Also, as the paper version of the measure can be used in the absence of an IT set-up and fulfils different administrative needs and will hence reduce administrative burden.

Study 1. Clinical validation of the M&MS questionnaire

Background

Measurement of mental health in children to date has typically been achieved by measures completed by proxies. Clinical settings in particular often rely on parent or clinician reported symptoms. With the increasing focus on children's perspective being important and necessary ("Children Act," 2004; "UN Convention on the Rights of the Child," 1989) which is reflected in policy focus on shared decision making in health services and the concept of self-defined recovery (Kennedy, 2010) there is a real need for measures that are valid and reliable for younger children. Moreover, research indicates that that children as young as 7-8 years are able reporters of their own mental health (e.g., Chorpita, Yim, Moffitt, Umemoto, & Francis, 2000; Sharp et al., 2006) . In community settings, particularly schools, self-report measurement supports screening for problems and early intervention (Levitt et al., 2007). In a clinic setting, this corresponds to, Patient Reported Outcome Measures (PROMs), which allow young people themselves to report on their health status. PROMs have gained importance and are more recently being highlighted as good practice in both policy and academic literature (Kennedy, 2010; Schmidt et al., 2001). Hence, the development of the M&MS (Deighton et al., 2013) filled a necessary gap for a free, simple self-report screening measure of child mental health that was suitable to use for younger populations and covers both emotional and behavioural difficulties.

The current study

The present study aims to test the ability of the measure to discriminate between a clinic and community sample (discriminant validity), examine the internal consistency of the measure in a clinical sample (internal reliability), assess the sensitivity of the scales to diagnoses (interpretability), compare it to another self-report measure (construct validity) and investigate agreement with parent completed questionnaires.

Method

Participants

Clinic Sample. Data were collected from 91 (46.2% female, N=42) young people attending two community teams from child and adolescent mental health services in an urban location in England: one generic team supporting the local area (67%) and one specialist team for vulnerable children (looked after, adopted, offenders; 33%). Ages ranged from 8 to 15 years (M=12.34, SD=2.03). A large proportion of the sample belonged to the White British ethnic group (69%, N=63) and the remaining participants were Asian (N=8), Mixed (N=6) and 15% (N=14) did not have a recorded ethnicity on file.

Participants completed the questionnaire either before or after their session with the clinician in the mental health service. Parents/carers and young people were given information about the study and asked for their consent. Participants were informed of the confidentiality of their responses and their right to decline to participate.

Community sample. For comparative analysis with a community sample matched controls were selected from a sample of young people who had completed the questionnaire in a school setting. Data collected in the same year as part of a school based study from 863 students from 7 schools in urban locations were used to find matched controls (4 primary and 3 secondary schools, ages 8-15 years, M_{age}= 11.97, SD=1.65; Female 48.9%, N=422; Ethnicity 63.6% White).

Surveys were completed in classroom-based sessions facilitated by researchers. Consent was sought from parents via mail beforehand. All participants received information about the study, including explanation of the confidentiality of their responses and their right to decline to participate and drop out at any time.

The community sample was matched to the clinic sample to control for differences in samples biasing the results. This was done because risk of mental health problems has shown to be varied based on gender, ethnicity and age (Green et al., 2005). The matched sample from the larger pool of controls was selected using propensity score matching (Rosenbaum & Rubin, 1985). Propensity score matching is an especially useful method when case-control matching on many variables as it provides a natural weighting scheme based on which propensity scores are created and these scores are used to select a matched sample. In propensity score matching as the number of variables to be matched on increases, and/or if variables are polytomous or continuous (rather than binary) it becomes more difficult to find exact matched controls for each individual in the case group (Dehejia & Wahba, 2002). Matching was carried out using *psmatch2* (Leuven & Sianesi, 2003) in STATA (StataCorp, 2011) which resulted in a matched community sample of 91 participants (49.5 % females (N=45), 68.6% White British, M_{age}= 12.29, SD=1.87). Table 1.1 includes descriptives and comparison between the two samples on key demographic variables and difference tests indicate that the two samples did not differ based on these demographic variables.

Table 1.1. Showing demographic characteristics of the clinic and community sample

Demographic	**Clinic**	**Community**	**Difference test**
Gender (% Female)	46.2%	49.5%	χ^2=0.20, p=.66
Age M (SD)	12.35 (2.02)	12.29 (1.87)	t(180)= 0.22, p=.83
Ethnicity (% White)	69.2%	62.6%	χ^2 =0.88, p=.35

Measures

Me and My School (M&MS). The M&MS questionnaire (Deighton, Tymms, et al., 2013) is a 16-item measure comprising of a 10-item emotional difficulties scale and a 6-item behavioural difficulties scale. Items in the emotional difficulties scale include 'I feel lonely'

and ‘I worry a lot’; items in the behavioural difficulties scale include ‘I lose my temper’ and ‘I hit out when I’m angry’ (Full measure in Appendix B). Participants respond to each item by selecting one of three options: Never, Sometimes, Always. Total scale scores are created by summing the item scores which results in a possible range of scores of 0-20 for the emotional and 0-12 for the behavioural difficulties scales, a higher score indicating more problems. In case of missing items prorated imputation is conducted up to missing 3-items for the emotional and up to 2-items for the behavioural scale. During the validation of the measure cut-off scores with clinical significance were established resulting in a score of 10 and above indicating problems on the emotional difficulties scale (10-11 borderline, 12 + clinical) and 6 and above indicating behavioural problems on the behavioural difficulties scale (6 borderline, 7+ clinical). The paper-pencil measure was used in the present study.

Strengths and Difficulties Questionnaire (SDQ) self-report. The SDQ self-report (Goodman, Meltzer & Bailey, 1998) is a self-report measure of mental health suitable for children older than 11 years. The measure consists of five 5-item scales, Emotional Symptoms, Conduct Problems, Hyperactivity, Peer Problems and Prosocial. The first four scales also sum to give a total difficulties score. This questionnaire was completed by 56 participants (57% female, N=32) in the clinical sample who were old enough (M_{age} = 13.46, SD = 1.29).

Parent SDQ. Accompanying parents or carers were also asked to complete the parent version of the Strengths and Difficulties Questionnaire which like the self-report version is a 25 item measure with five scales (Goodman, 1997). 92% (N=84) of accompanying parents/carers completed the questionnaire (58.3% mothers, 10.7% fathers, 3.6% other and 28% not known).

Clinical diagnoses. ICD-10 clinical diagnoses that were assigned to the young people attending mental health services were recorded. Recorded diagnoses could take the form of

ICD-10 diagnostic criteria or when no diagnosis was present, a Z-code which represents reasons for treatment (examples: lack of warmth in parent-child relationship (or) disability). For individuals with no diagnoses, under assessment or a Z-code, presenting problems were recorded. Only 54% (N=49) had a diagnosis and the rest of the sample did not have any diagnoses assigned. Out of the 42 participants who did not have a diagnosis only 26 had some kind of presenting problem recorded. Presenting problems might have been recorded by referring school, clinician, parent or social services.

Two clinical child psychologists then independently classified the diagnoses and presenting problems into groupings based on their clinical expertise and experience. The groupings used were *emotional, behavioural, emotional and behavioural* and *other*. This classification resulted in a complete agreement in coding for 82% of the items and any disagreements between the two coding clinicians were resolved in a discussion to ensure there was a clear classifying system (final list of the categorisation is in Appendix C). Based on this classifying system, for example, depression and anxiety were classified as *emotional* and learning disorders, hyperactivity, autism, and tourette's were in the *Other* category. These groupings were then applied to assign participants' diagnoses (and in the absence of a diagnosis, their presenting problems) to these groups. This resulted in 34 individuals with *emotional*, 7 individuals with *behavioural*, 13 individuals with co-morbid *emotional and behavioural* and 25 individuals in the *other* clinical assessment grouping.

Analysis and Results

Analyses were carried out in four stages to specifically look at different psychometric properties of this measure. In the first stage, internal consistencies were computed to assess reliability of the scale in the clinic setting. In stage two the ability of the M&MS to discriminate between clinical and community samples was assessed using mean comparisons, receiver operating curves (ROC) and comparing proportions above the scales' clinical thresholds. In the third stage the predictive validity of the *emotional difficulties* and *behavioural difficulties* scales was examined using clinical assessment. Lastly, correlations between the M&MS and Parent SDQ and SDQ self-report were explored to assess inter-rater reliability and construct validity.

Internal Reliability

Cronbach's alpha for the two M&MS sub-scales in the clinical sample were good: emotional difficulties, α=.84, behavioural difficulties, α=.82. The reliabilities in the community sample were slightly lower: emotional difficulties, α=.77, behavioural difficulties, α=.77, which is similar to the internal reliabilities obtained in the community sample in the initial validation (Deighton et al., 2013). Comparatively, in the clinic sample, the internal reliabilities were slightly lower for both the self-report SDQ (emotional symptoms, α=.83, conduct problems, α=.75) and parent completed SDQ (emotional symptoms, α=.80, conduct problems, α=.76).

Discriminating between clinic and community samples

As can be seen in Table 1.2 mean scores on both the scales were significantly higher in the clinic sample when compared to the community sample. For the emotional difficulties scale, on average there was a difference of more than 4-points on the scale ($t(167.49) = -7.87$, $p<0.001$, $d=1.17$) and for the behavioural difficulties scale an average difference of 2.7 points on the scale ($t(166.95) = -7.58$, $p<0.001$, $d=1.12$).

Table 1.2. Comparisons between the Clinic and Community samples for the Emotional and Behavioural Difficulties Scales of the M&MS

Scale	Sample	Mean (SD)	Mean Comparisons	Area Under the Curve (95% CI)
Emotional difficulties	Clinic	8.65 (4.06)	-7.87; p< 0.001, d=1.17	.79 (.73-.86)
	Community	4.40 (3.14)		
Behavioural difficulties	Clinic	5.13 (2.74)	-7.58 p< 0.001, d=1.12	.78 (.71-.84)
	Community	2.42 (2.05)		

To estimate the sensitivity of the measure and its ability to discriminate between the community and clinical sample ROC analysis was conducted for both scales. ROC curves are based on statistical decision theory and demonstrate the ability of a test to discriminate between alternative states of health (Zweig & Campbell, 1993), in this case mental health. The main statistic, the Area under the Curve (AUC), represents the probability that the measure will discriminate a positive (clinical/ at-risk) case from a negative (community/ low-risk) case. The AUC statistic for the *emotional difficulties* scale was .79 (SE= .03; 95% CI .73-.86) and for the *behavioural difficulties* scale was .78 (SE= .03; 95% CI .71-.84).

In terms of participants having scores higher than the threshold score for problems, on the *emotional difficulties* scale 40% of the clinic sample had high scores whereas 8.8% of the community sample had high scores. On the *behavioural difficulties* scale 41% of the clinical sample had above threshold scores as compared to 6.6% of the community sample. Overall, 58% of the clinic sample and 12% of the community sample had an above threshold score in either scale which represents overall sensitivity of the measure to individuals with risk.

Sensitivity of the scales to diagnoses

The sensitivity of the two scales to clinical diagnoses was explored in this section. As sample numbers are small for each diagnostic grouping, a descriptive approach was used, and is only meant as a way of illustrating the clinical utility and interpretability of the two-

subscales. As can be seen in Table 1.3, the emotional difficulties scores showed sensitivity to diagnoses and presenting problems. For instance, for participants with emotional related diagnoses the mean *emotional difficulties* score was 10.76 (SD= 4.13). The behavioural difficulties scale also showed sensitivity to clinical assessment – individuals in the *emotional/behavioural* group had a mean (SD) of 8.08 (4.95) The cut-offs on the scales also indicate sensitivity to diagnoses and presenting problems. The proportion of participants who had a diagnosis in the emotional domain and had high scores on the *emotional difficulties* scale was 88%, similarly 60% of those identified with behavioural presenting problems had above-threshold scores on the *behavioural difficulties* scale.

Table 1.3 Emotional Difficulties and Behavioural Difficulties Scales by Clinical Assessment

Clinical Assessment Grouping (N)	**Emotional difficulties scale**		**Behavioural difficulties scale**	
	Mean (SD)	% above threshold (N)	Mean (SD)	% above threshold (N)
Emotional (34)	10.76 (4.13)	65 (22)	5.16 (2.48)	41 (14)
Behavioural (7)	4.67 (2.65)	0 (0)	4.57 (2.23)	29 (2)
Emotional/Behavioural(13)	8.08 (4.95)	31 (4)	7.00 (2.61)	77 (10)
Other (25)	7.76 (3.43)	32 (8)	4.61 (2.99)	32 (8)
Z-code (30)	8.09 (3.85)	37 (11)	4.77 (2.73)	43 (13)
No Diagnoses (7)	7.00 (2.61)	14 (1)	6.49 (2.20)	57 (4)

Correlations with Parent SDQ and SDQ-self-report

Table 1.4 presents the correlations between the emotional and behavioural scales of the M&MS, Parent SDQ and SDQ self-report.

Table 1.4. Correlations between M&MS, Parent SDQ and SDQ- self-report

Variable	1.	2.	3.	4.	5.
1.M&MS emotional difficulties	-				
2.M&MS behavioural difficulties	.12	-			
3.Parent SDQ emotional symptoms	.3**	.1	-		
4.Parent SDQ conduct problems	-.27*	.3**	.08	-	
5.SDQ self-report emotional symptoms	.85***	.11	.41**	-.17	-
6.SDQ self-report conduct problems	-.07	.56***	.01	.46**	-.18

*p<.05; **p<.01, ***p<.001; *Note.* Sample size for M&MS – Parent SDQ assessments N=82-83, M&MS – SDQ-SR N=52-53, Parent SDQ – SDQ-SR N=48

The correlations between the corresponding scales of the parent SDQ and the M&MS were both 0.3 and significant at p<0.001. In terms of correlation with the self-report version of the SDQ, completed by only 11+ year old participants, the corresponding emotional scales correlated highly r=0.85, and the behaviour scales had moderately high correlations (r=0.56). The non-corresponding scales had very low correlations (0.11 & -0.07), which are comparable to the correlations between non-corresponding scales of the M&MS (r=0.12) and SDQ-self-report (r= –0.18).

Discussion

Analyses indicate that both the scales, emotional difficulties and behavioural difficulties of the M&MS questionnaire sufficiently discriminate between a clinic (at-risk) and community (low-risk) sample. The amount of discrimination as represented by the AUC statistics (emotional difficulties=0.79, behavioural difficulties= 0.77) are comparable to the AUC of the emotional symptoms scale (0.75) and the conduct problems scale (0.77) of the self-report SDQ (Goodman et al.,. 1998). Mean differences were statistically significant with large effect sizes (d> 1.1) with individuals in the clinic sample being 4.5 times more likely to be above the threshold indicating problems on the emotional difficulties and 6 times more likely to be above the threshold indicating problems on the behavioural difficulties scale. Sensitivity of the individual scales was 40-41%, in comparison the scale sensitivity is 22-32% for the scales of the youth self-report (YSR) of the ASEBA (Achenbach & Rescorla, 2001). The overall sensitivity of the scales' thresholds was 58%, which is comparable to the 59% found for the SDQ (Goodman et al., 1998). Comparatively, on the 105-item YSR overall sensitivity was 65% (Achenbach & Rescorla, 2001) which considering the greater length and diagnostic focus of the YSR can be expected. This suggests that even though the M&MS is brief measure with a general mental health focus it captures clinical need to a reasonable level. In terms of the community sample 12% had scores above the threshold which is lower than the 23% found for the SDQ but is more in accordance with the 9-12% expected from a normal representative population in this age range (Ford, Goodman, & Meltzer, 2003; Green et al., 2005).

Overall, though the sensitivity of the measure is comparable to other self–report measures of mental health, it is not high. The reason for generally low sensitivity for child mental health measures could be partly due to the lack of diagnosis for a large part of the sample as mental health services see a wide range of problems that might be related to the family situation or care provision; which though serious do not always lead to a diagnosis.

This is supported from studies of service use which demonstrate that having a diagnosis does not increase the likelihood of speciality mental health care versus use of other services (Angold et al., 2002) and that a substantial proportion of young people who attend mental health services have no diagnosis (Burns et al., 1995). Additionally the community sample includes individuals with either diagnosed or sub-clinical levels of problems as the sample was not filtered to only include 'healthy' individuals. Higher sensitivity on these measures might be expected if the clinic sample were drawn from an intensive or in-patient psychiatric unit and/or the community sample did not include any individuals with problems.

Cross-referencing clinical assessment with the scale scores provides some evidence that both scales are responsive to clinical diagnoses as indicated by the mean scores and proportion above the clinical threshold in each diagnostic group. Individuals with emotional related problems had high scores on the *emotional difficulties scale* with more than 65% having scores above the clinical threshold. In terms of the *behavioural difficulties scale* there was a discrepancy between individuals assessed as having just behavioural symptoms and those with co-morbid emotional and behavioural symptoms in terms of their reporting of behavioural symptoms. A much smaller proportion of those assessed as having only behavioural problems scored above the threshold on the behavioural difficulties scale (29%) in comparison to those with comorbid emotional and behavioural problems (77%). Given the numbers are small this could be a chance observation but the finding suggests that children with only behavioural assessment might have more difficulties perceiving problems with their own behaviour. Additional research is required specifically exploring this discrepancy in self-reporting behavioural problems, as externalising problems might be linked with lower self-awareness of these problems, thereby affecting self-reports (Yammarino & Atwater, 1993)

The measure has good internal reliability as indicated by the Cronbach's alphas of the two sub-scales (*emotional difficulties*, α=.84; *behavioural difficulties*, α=.82). The correlations

between the corresponding scales of the M&MS and self-rated SDQ were high (*emotional difficulties* r=.85; *behavioural difficulties*, r=.56) and compared favourably to the correlations found in community samples (*emotional difficulties* r=.67; *behavioural difficulties* r=.70; Deighton et al., 2013) which supports the construct validity of the measure in the clinic setting. Correlations between the corresponding scales self-reported measures and parent SDQ were similar for M&MS-Parent SDQ (.3) and the SDQ-Parent SDQ (.4). Overall the inter-rater correlations for the M&MS were in line with expected correlations between parent and child report which are generally not very high (Achenbach et al., 1987; Kolko & Kazdin, 1993) and were comparable to results from other measures (average r= .25) found in a meta-analysis (Achenbach et al., 1987).

While clinical assessments provide early indication of the scales' sensitivity to case type, the small numbers identified within each diagnostic category mean that formal statistical testing could not be carried out. This is something that could be explored in further studies as the measure is used more widely with clinic samples. Of particular interest is the consideration of an amendment to the clinical thresholds of the *emotional difficulties* scale. In the initial validation (Deighton et al., 2013), thresholds were computed using an equi-percentile approach in the community sample. The results of this study indicate that a lower threshold might capture clinical levels of emotional problems better. Future research in clinic samples should explore this possibility to ensure the measure has optimum screening capability.

The primary aim of the current study was to establish the credentials of M&MS as a screening tool for use in community settings. As such, results indicate that the measure discriminates sufficiently between clinic and community samples. However, two areas require development if the measure is to be used more widely used in clinic settings: the scales' sensitivity to case type or diagnosis and its responsiveness to change over time.

In clinical settings, a lot of focus is being placed on user satisfaction and patient reported outcomes. However to date this has been more problematic for younger populations, with almost all data being provided by adult proxies on their behalf. The development of this measure has the potential to fill this gap for a self-report measure for pre-adolescents in clinical settings. Moreover, using the same measure in community and clinical settings will aid consistency and transferability and can improve interpretability across settings. In addition, considering a dearth of self-report behavioural scales that can be used for outcome monitoring in services, it is also being used as part of a UK national initiative to improve quality of psychological therapies and a programme promoting use of outcome monitoring in child and adolescent mental health services in the UK. However, further research in clinic settings will be necessary to establish the scales sensitivity to diagnosis and responsiveness to change over time.

In conclusion, the findings of this study indicate that this measure sufficiently discriminates between at-risk (clinic) and low-risk (community) samples, has good internal reliability, compares favourably with existing self-report measures of mental health and has comparable levels of agreement between parent-report and self-report to other measures. Alongside existing validation of the M&MS (Deighton et al. 2013), these findings justify the measure's use as a self-report screening tool for mental health problems in community settings, and lay out its potential as a patient-reported outcome tool in clinic settings.

Study 2. Psychometric equivalence of the computer and paper survey formats of the M&MS questionnaire

Background

Computers are increasingly being used with adult and child populations to complete questionnaires, whether for population-based epidemiological surveys, assessing health outcomes in services or screening for problems. Computer-based survey methods are recognised as having many benefits over paper survey methods, such as increased efficiency of data collection and management and reduced coding errors, which in turn increase the speed at which feedback and results can be produced (Hayslett & Wildemuth, 2004; Kays, Gathercoal, & Buhrow, 2012). Even though computer based surveys have many advantages, paper based surveys may sometimes be preferable, especially in clinical settings and in settings where access to computers is limited.

Survey format effects

However, studies of questionnaire format have found that survey format influences survey response rates (Hayslett & Wildemuth, 2004), item response and missed items, especially for items of a sensitive nature (Kays et al., 2012), and social desirability effects (Booth-Kewley, Larson, & Miyoshi, 2007). This indicates psychometric equivalence between different survey formats such as paper based and computer based cannot be assumed. In support of this, the American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999) highlighted the need for cross-format equivalences to be established prior to direct comparison of data collected from paper-pencil based surveys and computer and internet based surveys. The Scientific Advisory Committee of the Medical Outcomes Trust (2002) included validation of alternate forms of measures into their measure assessment framework for measures of health outcomes and quality of life.

While psychometric equivalence is not routinely tested in mental health measurement, especially in child and adolescent mental health, some widely used adult mental health

measures have been investigated, yielding mixed results. Holländare, Andersson, & Engström (2010) tested equivalence of the Becks Depression Inventory (BDI-II) and the Montgomery-Asberg Depression Rating Scale- self rated (MADRAS-S) and found partial format effects for the BDI-II. In this study they tested equivalence using the same participants and as a result also found order effects (paper first had higher scores) and order and format interaction effects as well (paper-first group scored significantly higher on the paper BDI-II than on the computer based version). Wijndaele et al. (2007) assessed the equivalence of five mental health related measures in adults, including the General Health Questionnaire (GHQ-12) and Symptom Checklist (SCL-90-R). Equivalence varied depending on the measure with low test-retest co-efficients for the SCL-90-R and high co-efficients for the GHQ-12 and MOSSS. Whitehead (2011) tested equivalence of the Hospital Anxiety and Depression Scale (HADS) and the Fatigue Symptom Inventory (FSI) using separate samples and found significantly higher fatigue being reported online. Hence, it is critical to establish when and where there is cross-format equivalence before data collected from multiple formats can be considered comparable.

While different formats of many of the most widely used child self-report mental health measures exist, such as the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) and the Achenbach System of Behaviour Assessment (ASEBA; Achenbach & Rescorla, 2001), equivalence testing is not common practice. In order to justify equivalence of these different formats for children, it is not sufficient to generalise from findings from adult measures. There are many differences in how young people experience computer-based environments compared with adults (Prensky, 2001). Current generations of young people are not only more comfortable and proficient with computers and newer technologies; but they also use them to self-express and are more comfortable disclosing sensitive information on these media (Livingstone, 2008; Turner et al., 1998). The proposed explanation is that young people's conception of privacy and intimacy when using computers and the internet are

different from that of previous generations, mainly as a result of having grown up with access to these technologies (Helsper & Eynon, 2009; Livingstone, 2008). This phenomena where individuals demonstrate greater disclosure online that they might have in person or not-online has been termed the 'online disinhibition effect'(Suler, 2006).

Psychometric equivalence of measures in different survey formats, therefore, needs to be tested and not assumed. The few measures of mental health that have been tested for equivalence are adult measures, not child measures, and there is reason to believe that results from adults might not apply to young people today as they are more comfortable using computer based mediums.

Methodological approach

Existing research testing equivalence between survey formats consistently uses scale-level approaches such as means, internal consistency, correlations, and testing constancy of factor structure across formats. These statistics are necessary but not sufficient to establish equivalence of survey formats as they provide information only at the scale-level. Scale-level analyses do not account for how individuals at different levels of the latent construct perform on the individual items of the instrument (Hambleton & Jones, 1993; Raykov & Marcolides, 2011). This limitation is dealt with in item-level approaches, such as IRT, as at the core of these approaches is a model that describes how individual subject responses on items of an instrument relate to an unobservable trait (Hambleton & Jones, 1993; Raykov & Marcolides, 2011). In other words IRT approaches model how the probability of a response to an item (in mental health this is usually equivalent to endorsing the existence of a symptom) varies as a function of the individual's location on the latent continuum or trait (Santor, Ramsay, & Zuroff, 1994).

One of the key constructs within the item-based framework is looking at item response probability based on different groups, which is termed differential item functioning (DIF;

Walker, 2011). DIF occurs when individuals at the same level of the trait or construct being measured have unequal probabilities of attaining a given score on a given item, usually based on socio-demographic grouping such as gender or ethnicity (Rogers, 2005). DIF analysis therefore attempts to disentangle these item-performance differences while controlling for overall score on the latent trait. DIF analysis is conducted with 2 groups (one is called the focal group and the other the reference group) and traditionally these groupings have been demographic such as gender or ethnicity based groupings (Rogers, 2005). It has been recognised that it is important that equivalent scores obtained, for instance, from a measure of depression reflect the same construct for all respondents regardless of gender or ethnicity (Santor et al., 1994; Walker, 2011).

In mental health measurement, especially child and adolescent mental health measurement, IRT techniques such as DIF analysis are not yet widely used (Sharp et al., 2006) even though they have been around for a few decades now and are routinely used in measure construction and evaluation in other fields, especially education. The reasons for this have been attributed to several factors, mainly that researchers in these fields are unaware of these newer methodologies and the advantages they offer over traditional CTT methods (van der Linden & Hambleton, 1997) and hence these methods have had less coverage in journals (Embretson & Reise, 2000). More recently measures of child and adolescent mental health such as the Short Moods and Feelings Questionnaire (SMFQ, Sharp et al., 2006), the Kidscreen measures (Ravens-Sieberer, Gosch, et al., 2007) and the M&MS (Deighton, Tymms, et al., 2013) have used IRT techniques either to justify existing scale properties and items (Sharp et al., 2006) or to help in item selection (Deighton, Tymms, et al., 2013; Ravens-Sieberer, Gosch, et al., 2007). Petersen et al. (2003) used DIF analysis to compare translations of an emotional functional scale (EORTC QLQ-C30). Extending this use of the IRT framework to assess for Differential Item Functioning across different groups, the

methodology can be suitably applied to assess whether there are differences in item functionality across different survey formats (paper and computer) of the same instruments.

The current study

In light of the mixed results found with adult measures of mental health and literature highlighting the differences between current generations of young people and adults in terms of the relationship they have with computers and the internet, we propose to test the psychometric equivalence of the computer and paper survey formats of the M&MS questionnaire. This is the first study, to my knowledge, looking at equivalence between the paper and computer based formats of a child and adolescent self-report measure of general mental health.

In terms of study methodology used to test psychometric equivalence between survey formats, there are two approaches that are commonly used; same sample repeated measurement (e.g., Holländare et al., 2010) and comparison of separate demographically similar samples (e.g., Ritter et al.,2004). In this study I decided to use demographically matched samples to avoid practice and order effects associated with repeated measurement (as found in Hollander et al., 2010). This approach is also supported by the finding that score instabilities and lower test-retest reliabilities for self-reported mental health in young people are more common (e.g., Goodman, Meltzer & Bailey, 1998) as children might be less consistent in their self-perception regarding difficulties and might not respond based on stable adjustment but rather the 'here and now' (Wolpert et al., 2008).

This study seeks to establish if there is psychometric equivalence between the paper and computer based versions of a child self-report mental health measure the 'Me and My School (M&MS)', both at the scale level and at the item level, with the expectation that given children might be more comfortable disclosing information on a computer based medium there might be psychometric in-equivalences.

Method

Participants

Participants were school aged children in England from school Years 4 to 9 (ages 8-14 years) in the English school system. Socio-demographic information collected included gender, ethnicity and Free School Meal eligibility (FSM), which is often used as a proxy for socio-economic deprivation in school based research in England (Hobbs & Vignoles, 2010).

Paper surveys. Paper survey data were collected from three secondary schools and four primary schools across England. A total of 863 pupils from seven schools answered the paper version of the Me and My School questionnaire (M&MS). Out of these 777 (90%) participants had completed all the items from the behavioural difficulties and emotional difficulties scales. For item level analysis all items need to be complete and for this reason only pupils who completed all items were included in the analysis.

Computer surveys. The computer based survey comparison group were selected from a large comparison pool of 39,168 pupils (comprising 87.3% of the total sample with no missing items) from 630 primary and 180 secondary schools who completed the survey as part of a national study of mental health in schools (Wolpert et al., 2011).

A one-one matched community sample was created using propensity score matching, which allows matching based on many criteria simultaneously. A matched control group was selected by matching on year group, gender, ethnicity and free school meal eligibility to create a one-to-one matched comparison group. This method is akin to a random allocation approach and ensures that any differences between the two groups are not due to differences in key demographic predictors (such as age, gender and ethnicity; Bland & Altman, 1994). Similar to the previous study the matching was carried out using *psmatch2* (Leuven & Sianesi, 2003) in STATA (StataCorp, 2011) to create a one-to-one matched comparison group. Due to the large size of the control pool exact one-to-one matches were found for all

individuals in the paper survey sample. For instance a 10-year old, White male eligible for free schools meals who completed a paper survey was matched exactly to another 10-year old White male eligible for free school meals who completed the computer survey.

After taking into account demographic variables (gender, age, SES and ethnicity) the sub-sample of 777 computer based surveys were not different from, and were hence representative of, the large pool of 39,168 pupils that did the computer based survey (emotional difficulties, $\beta=0.002$, $p=0.72$; behavioural difficulties, $\beta=0.005$, $p=0.35$).

In each group (paper and computer, N=777) participants belonged to Years 4 to 9 (7.3% Y4, 15.4% Y5, 7.1% Y6, 11.8% Y7, 33.7% Y8, 24.6% Y9). 51.7% of the sample was male (N=402). Participants' ethnicity was grouped into six categories based on the broader groupings of ethnicity in the National Pupil Database in England. 53.7% of the sample was classified as White British, followed by 16.9% Black, 10.4% White Other, 9%Mixed, 7.1% Asian, 1.9% Any Other Ethnic Group and 1% of participating pupils were Unclassified. 23% (N=179) of the sample were eligible for Free School Meals. In terms of representativeness, overall the sample was more deprived than national school population (sample 23% eligible for FSM; national 12-14%) and had a lower proportion of pupils classified as being White British compared to nationally (sample 53.7%; national 73-77%; Department for Education, 2010b).

Procedure

Paper surveys were completed in classroom-based sessions facilitated by researchers. Computer-based surveys were completed by pupils on computers in school within the normal school day with support from their class teachers. Pupils could access their questionnaire with a unique code that was assigned to them. Both the online and paper versions presented items in a clear and child-friendly manner and in exactly the same sequence, the key difference

between the formats is that in the computer version items were presented one at a time on the screen whereas in the paper version they were presented one below the other.

For all data collection (in both survey formats) consent was sought from parents via post beforehand. All pupils received information about the study including explanation of the confidentiality of their responses and their right to decline to participate and drop out at any time. This information was then re-iterated by either the teacher (computer version) or the researcher (paper version) prior to pupils completing the survey.

Measures

Me and My School questionnaire

The Me and My School Measure questionnaire (M&MS; Deighton et al., 2013) has been described in the previous study and a complete list of items is available in Appendix B. DIF analysis informed the selection of the items in the measure and the final items in the measure do not exhibit DIF based on demographic groupings such as English as an additional language, special educational needs and socio-economic status.

The measure has at-risk thresholds with a score of 10 and above (10-11 borderline, 12 + clinical) indicating problems on the emotional difficulties scale and 6 and above (6 borderline, 7+clinical) indicating behavioural problems on the behavioural difficulties scale (Deighton et al., 2013). Given the focus on item-level analysis in the current paper Table 2.1 presents item response descriptives for the paper and computer survey samples.

Table 2.1. Item response proportions in the paper and computer survey formats

	Item response %					
	Paper survey			Computer survey		
Item	Never	Sometimes	Always	Never	Sometimes	Always
Emotional difficulties scale						

I feel lonely	64.7	33.7	1.5	61.1	34.7	4.1
I cry a lot	69.0	29.5	1.5	65.1	30.6	4.2
I am unhappy	40.9	57.5	1.5	38.0	58.9	3.1
Nobody likes me	64.1	33.1	2.8	66.7	28.3	5.0
I worry a lot	39.9	53.7	6.4	37.3	52.8	9.9
I have problems sleeping	65.6	28.7	5.7	58.7	31.5	9.8
I wake up in the night	41.6	50.5	8.0	36.9	49.3	13.8
I am shy	41.6	51.4	7.1	38.6	52.6	8.8
I feel scared	65.4	33.5	1.2	57.4	39.4	3.2
I worry when I am at school	68.2	28.4	3.3	61.1	33.5	5.4
Behavioural difficulties scale						
I get very angry	42.3	50.2	7.5	35.9	51.7	12.4
I lose my temper	46.6	45.8	7.6	40.2	48.2	11.7
I hit out when I am angry	64.7	27.7	7.6	55.1	33.2	11.7
I do things to hurt people	79.2	19.6	1.3	73.0	23.2	3.9
I am calm	5.4	60.2	34.4	7.3	60.6	32
I break things on purpose	87.9	11.2	0.9	80.7	14.9	4.4

Demographic variables

Gender. Information regarding pupils' gender was received from the school lists and national pupil database (NPD). Additionally pupils also indicated whether they are male or female on their questionnaire.

Ethnicity. Information about pupil's ethnicity was received from school records and detailed ethnic categories were grouped into the NPD broad groupings- White British, White other, Asian, Black, Mixed, any other ethnic group and unclassified.

Free school meal (FSM) Eligibility. Information on FSM eligibility status was received from the schools and is a part of the NPD.

Analysis and Results

Various analyses were carried out to establish the amount of equivalence between the paper and computer based versions of the measure. Five steps were taken to establish the level of equivalence between the paper and computer-based versions of the measure. First, scale-level mean comparisons were carried out for the emotional and behavioural difficulties scales reported via paper and computer-based formats. Second, a categorical data confirmatory factor analysis was conducted to confirm whether the known existing factor structure of the measure fitted the data collected from paper questionnaires. Third, internal reliabilities of the scales in both formats were compared using the Cronbach's alpha coefficient of internal consistency. Fourth, DIF analysis was conducted to assess the equivalence of item-response probabilities in the two formats. Lastly, Differential Test Functioning (DTF) analysis was conducted to compare how the entire set of items functioned in the different survey formats

Means and variances

Means and standard deviations were computed for the paper survey and computer survey samples and are shown in Table 2.2 The analysis was first conducted for the entire sample and was subsequently split by age into the primary and secondary school sample to assess if sample age influenced the results. Overall, as can be seen in Table 2.2, mean scores were significantly lower in the paper survey sample compared with the computer survey sample for both emotional difficulties (effect size, $d = 0.2$) and behavioural difficulties ($d = 0.24$). Owing to the large age range in the current sample, further analyses were conducted separately in the primary school aged children (8-11 years) and the secondary school aged adolescents (11-14 years) to assess if age-specific differences were present. This division by age is also of interest because, as mentioned in the introduction, most self-report measures of mental health have been designed and validated for children aged 11 years and above. As can be seen from Table 2.2, the format based differences were found in both the younger and older age groups.

Table 2.2. Means and Standard Deviations of emotional difficulties and behavioural difficulties scales for the overall sample and sub-samples broken down by age

		Overall sample		***8-11 years***		***11-14 years***	
Scale	Statistic	Paper survey	Computer survey	Paper survey	Computer survey	Paper survey	Computer survey
Emotional Difficulties Scale	M (SD)	4.78(3.23)	5.4(3.58)	5.21(3.37)	6.17(3.31)	4.60(3.15)	5.16(3.64)
	t-test (df)	t(1552) = 3.93***		t(462)=3.08**		t(1065.88)=2.74**	
	% above threshold	9.3%	13%	12.1%	16.4%	8.1%	11.6%
Behavioural Difficulties Scale	M (SD)	2.75(2.35)	3.35(2.65)	2.23(2.25)	3.24(2.44)	2.97(2.36)	3.39(2.73)
	t-test (df)	t(1552) = 4.72***		t(462)=4.64***		t(1064.9)=2.68**	
	% above threshold	11.7%	19.4%	9.5%	19%	12.7%	19.6%

***p<.001, **p<.01

Overall, mean difficulties in the paper survey sample when compared to the computer survey sample were significantly lower for both emotional difficulties (paper survey, M=4.78, SD=3.23; computer survey, M=5.46, SD=3.58; t (1552) = 3.93, p<0.001; d=0.2) and behavioural difficulties (paper survey, M=2.75, SD=2.35; computer survey, M=3.35, SD=2.65; t (1552) = 4.72, p<0.001; d=0.24). The effect did not vary by age-group and the significant mean differences in both the scale scores were still present.

Additionally, to assess if the format effects varied depending on gender an ANOVA was estimated. Format, gender and a gender*format interaction were included in the ANOVA model predicting the scale scores. Significant effects were found by gender and format for both subscales in the expected direction (emotional difficulties: gender F (1, 1550) =56.60, p<.001, format F(1,1550)=15.97, p<.001; behavioural difficulties: gender F (1, 1550) =26.98, p<.001, format F(1,1550)=22.36, p<.001). Results for the interaction term for both scales indicate that the effect of survey format is not different based on gender (emotional difficulties: F(1,1550)=.204, p=.652; behavioural difficulties: F(1,1550)=.469, p=.494).

Confirmatory Factor Analysis

The M&MS school measure has two scales: emotional difficulties and behavioural difficulties and the factor structure of the scale is already known. A confirmatory factor analysis was conducted to ascertain if the known factor structure fit the data from paper completed surveys. As the items had likert-scale categorical responses a categorical item confirmatory factor analysis was conducted in Mplus6 (Muthén & Muthén, 2007).

The model was specified such that cross-loading between items and factors did not occur and the two scales were treated as two uni-dimensional scales. Factor loadings of the items in the two scales are greater than 0.3 (see Table 2.3 for factor loadings). The fit indices (CFI = 0.96; TLI = 0.95; RMSEA = 0.06) indicated good model fit based on the widely used criteria of model fit (Hu & Bentler, 1999). CFA co-efficients for the computer based survey

are also presented in Table 2.3 to allow for comparisons (CFI = 0.95; TLI = 0.95; RMSEA = 0.06).

Table 2.3. Standardised Loadings of the Confirmatory Factor Analysis

	Paper survey		**Computer survey**	
	CFA Factors		**CFA Factors**	
Item	I	II	I	II
Emotional difficulties scale				
I feel lonely	.67		.63	
I cry a lot	.69		.68	
I am unhappy	.66		.73	
Nobody likes me	.56		.64	
I worry a lot	.75		.68	
I have problems sleeping	.58		.62	
I wake up in the night	.50		.59	
I am shy	.34		.29	
I feel scared	.83		.72	
I worry when I am at school	.78		.76	
Behavioural difficulties scale				
I get very angry		.93		.87
I lose my temper		.94		.83
I do things to hurt people		.81		.85
I am calm		.71		.79
I hit out when I am angry		.62		.58
I break things on purpose		.60		.66

Internal reliability

Cronbach's alpha of the paper based surveys was .78 for the emotional difficulties scale and .81 for the behavioural difficulties scale and for the computer survey sample alpha was .80 for the emotional scale and .82 for the behavioural difficulties scale.

Differential item functioning (DIF)

DIF analysis was carried out to determine if any of the items behaved differently based on if the questionnaire was completed with paper-pencil or on a computer. DIF analysis

to determine if any of the items operated differently based on survey format can be conducted using a variety of approaches including IRT, Mantel- Haenszel, logistic regression and the Rasch model (Karami, 2012). Two common approaches were used here, 1) Liu–Agresti common log odds ratio (L-A-LOR; Liu & Agresti, 1996) which is based on the Mantel-Haenszel common-odds ratio generalised to polytomous data and represents the log odds ratio of one group selecting a response option compared with the other group when the level of the overall measured construct is the same, and 2) IRT approach with graded response model (Samejima, 1997) was used and the presence of DIF is indicated by the difference in model fit estimates on the $\boldsymbol{X^2}$ distribution between the model with both the item difficulty and discrimination parameters constrained to be equal and the model where they are allowed to be estimated freely. L-A-LOR was estimated in DIFAS 5.0 (Penfield, 2005) and IRT based DIF model was estimated in IRTPRO (Paek & Han, 2012). The size of the DIF was interpreted using a widely accepted classifying system whereby DIF in polytomous items is considered negligible if L-A-LOR < 0.43, moderate if between 0.43 and 0.64, and large if >0.64 (Penfield, 2007).

Table 2.4. DIF co-efficients, using both Mantel Haenszel and IRT based approaches

Item	**L-A-LOR (SE)** ***(focal group=Paper)***	**X^2difference (df=2)** ***(IRT approach)***
Emotional difficulties scale		
I feel lonely	-.01(.13)	4.9
I cry a lot	.01(.13)	4.4
I am unhappy	-.1(.13)	0.8
Nobody likes me	-.37*(.13)	12.9*
I worry a lot	-.14(.12)	2.2
I have problems sleeping	-.17(.12)	3.4
I wake up in the night	-.14(.11)	5.4
I am shy	.00(.12)	0.3
I feel scared	.17(.14)	1.8
I worry when I am at school	.09(.13)	0.3
Behavioural difficulties scale		
I get very angry	-.05(.14)	1.3
I lose my temper	-.07(.15)	0.7
I hit out when I am angry	.23(.15)	0.7
I do things to hurt people	.07(.26)	2.3
I am calm	-.37*(.13)	1.9
I break things on purpose	.38*(.17)	10.3*

* statistically significant at the 0.05 level. 2. Negative L-A-LOR values indicate DIF favouring the focal group i.e. for the same level of construct easier to endorse for the focal group. Conversely, positive L-A-LOR values indicate the item is more difficult to endorse for the focal group.

Table 2.4 presents the L-A-LOR and X^2difference values for all the items in the two scales. In terms of the L-A-LOR analysis, one item in the emotional difficulties scale, 'Nobody likes me', exhibited a statistically significant but negligible DIF based on the criteria outlined above (Penfield, 2007). Two items in the behavioural difficulties scale, 'I am calm' and 'I break things on purpose', also exhibited statistically significant yet negligible amounts

of DIF. Given the same level of the latent construct, positive L-A LOR values indicate the item is more difficult to endorse for the focal group; negative values indicate that the item is easier to endorse for the focal group. The negative L-A-LOR of 'I am calm' indicates that it was easier to endorse in the paper format, whereas the positive L-A-LOR value for 'I break things on purpose' indicates that it was easier to endorse in the computer based format of the measure. The results obtained via the IRT approach largely cross-validated the results found based on the L-A-LOR approach. The only discrepancy was that the IRT approach did not find significant DIF for 'I am calm' in the behavioural scale.

Additionally, DIF analysis by age-group (primary and secondary school aged participants, see Table 2.5 for co-efficients), indicated similar patterns in items that exhibited DIF in the two age-groups on the behavioural difficulties scale and one item 'I break things on purpose' exhibited significant and moderate DIF favouring endorsement in the computer format in older participants (L-A-LOR=.43). However, for the items in the emotional difficulties scale only younger participants exhibited significant and moderate DIF on the item 'Nobody likes me' (L-A-LOR= -.56), whereas older participants did not exhibit significant DIF on any of the items.

Table 2.5. DIF analysis by age group using the Mantel- Haenszel L-A-LOR approach

Item	**8-11 years** L-A-LOR (SE) *(focal group=Paper)*	**11-14 years** L-A-LOR (SE) *(focal group=Paper)*
Emotional difficulties scale		
I feel lonely	-.06(.22)	.04(.17)
I cry a lot	.01(.23)	-.04(.16)
I am unhappy	-.12(.23)	-.16(.15)
Nobody likes me	-.56*(.24)	-.29(.15)
I worry a lot	-.46(.24)	-.08(.14)
I have problems sleeping	.14(.21)	.18(.15)
I wake up in the night	.24(.20)	.12(.13)
I am shy	.08(.20)	-.04(.13)
I feel scared	.32(.23)	.10(.17)
I worry when I am at school	.01(.24)	.15(.16)
Behavioural difficulties scale		
I get very angry	-.13(.26)	-.02(.17)
I lose my temper	-.01(.27)	-.14(.18)
I hit out when I am angry	.29(.28)	.20(.18)
I do things to hurt people	.22(.30)	-.03(.19)
I am calm	-.45(.24)	-.32*(.16)
I break things on purpose	.47(.38)	.43*(.20)

* statistically significant at the 0.05 level

DTF

DTF assesses the aggregate effect of DIF across all the items in a scale (Penfield & Algina, 2006) and was analysed using the v^2 statistic in DIFAS 5.0 (Penfield, 2005). Co-efficients are presented in Table 2.6. Based on criteria for assessing the size of DTF (Penfield & Algina, 2006) a v^2 <.07 is considered negligible and hence the DTFs were deemed not to warrant concern.

Table 2.6.2 DTF analysis for the overall sample and the sub-samples by age group

	Emotional difficulties scale DTF v^2 (SE)	Behavioural difficulties scale DTF v^2 (SE)
Overall sample	.01 (.01)	.03 (.03)
8-11 years	.02 (.03)	.01 (.05)
11-14 years	-.003 (.01)	.03 (.03)

Discussion

Formal equivalence testing of different formats of measures used to assess child and adolescent mental health is not yet a widely adopted practice, and even though many of the most widely used measures are available in both paper and computer-based formats, not much is known about their equivalence across formats. The current study aimed to, first, test the psychometric equivalence of a child self-report mental health measure, 'Me and My School' and, secondly, demonstrate how DIF analysis can be used to assess item-level differences alongside current methodologies used to assess scale-level differences.

The results indicate that there are overall differences in mean scores for the paper and computer-based version of the M&MS questionnaire and the DIF analysis indicates that format differences at the item-level are almost non-existent, except for one item which displayed moderate DIF only in younger children. However, the DTF analysis in both the overall and age-specific samples suggested the effect of DIF across all the items was negligible. The discrepancy between the scale- level psychometric in-equivalences and the item-level equivalences suggest that the difference in scores between the formats is due to an overall 'dampening' of scores in the paper format. This might be attributed to differences in 1) level of disclosure to topics of a sensitive nature, such as mental health, in the different formats and 2) differences in perceived privacy and confidentiality afforded by the survey formats. Both of these points are related as they reflect the level of comfort and likelihood to disclose information based on the medium of survey, indicating that the increase in use of technology and social network sites by young people might have an influence on their readiness to disclose sensitive information via computer and internet based mediums (Livingstone, 2008; Turner et al., 1998), which has been termed the online disinhibition effect (Suler, 2006). Although our results so far indicate that young people might be more comfortable disclosing on sensitive issues like their mental health on computer based scales when compared to paper based ones, this difference can be further explored, in terms of

young people's levels of comfort with the different formats and the effects of different levels of format familiarity on item response. For instance, this can be done via qualitative studies of young people's views on privacy and confidentiality in the remit of these different media or via tests of physiological stress reactions to disclosing sensitive information in these different settings.

The DIF analyses carried out in the current paper illustrate how format-based differences can be assessed at item-level. The fact that the scale-level analysis and the item-level analysis lead to different conclusions about the equivalence of the M&MS measures emphasises the importance of using both methods. We suggest future studies looking to establish equivalence between formats also use this item level analysis alongside complete scale or measure level analyses to get a better understanding of where there are psychometric equivalences and in-equivalences between different survey formats of the same measure.

While the current study marks a step forward in methodological approaches to equivalence in child mental health measures, the main limitation is that pupils completing the questionnaires were not randomly allocated to the paper and computer survey conditions and the allocation to different formats was at the school level. While this is a consideration, research indicates that, in England, less than 3% variation in mental health scores is explained by the school level (Hale et al., 2013; Wolpert et al., 2011) which suggests that the results are not attributable to allocation at the school level. Although the proportions of pupils missing items in both the paper and the computer versions were similar, the current study does not explore in-depth the possibility of differences in missing items and their differential predictors in the two samples.

The results of this study raise the question of how to deal with psychometric in-equivalence across survey formats, when they exist. In the case of the M&MS as differences are at the scale level and not the item level it would be easy to account for format effects in

other analysis. Further research with this measure is required to assess if these results are replicated and if the amount by which the paper based surveys result in lower scores remains consistent in different samples and settings. As computer based survey administration is likely to only become more common this is an area of study that could benefit from more research and discussion.

While results of this study are not generalisable to other measures of mental health in young people, the difference in the scale scores between the two formats indicates that it may be necessary for other measures of mental health to test equivalence of formats before using different formats of measures widely and inter-changeably. The study raises concerns about how measures are currently used across all settings without sufficient tests of equivalence and strongly suggests the need for examining the impact of format when they are used simultaneously, especially in studies of intervention effectiveness. Even though the effect sizes of the difference for this measure would be considered small (Cohen, 1988), where non-random allocation of formats occurs, this could have a significant impact on outcomes of intervention studies. Moreover, when used for screening, the differences in proportions identified as being at-risk is large enough to warrant concern when used at a population level. Until further research is done to assess the psychometric equivalence of commonly used child mental health measures across formats, some degree of caution is warranted when combining or directly comparing data collected via different formats and in repeated measurement studies.

Part 1 Conclusions

The beginning sections of Part 1 systematically reviewed existing measures and assessed their properties in terms of criteria set out to evaluate measures, both in terms of psychometrics and practical utility. The results of this exercise supported the case for further validating the M&MS questionnaire as it fills a necessary gap for a child self-reported mental health measure that can be used by primary school-aged children and is free-to-use. The studies in this chapter aimed to further validate the M&MS questionnaire as a community based mental health measure and screening tool. The first study aimed to assess the discriminative capacity and interpretability of the scores of the measure by examining the ability of the measure to discriminate between at-risk and low-risk individuals. The second study aimed to assess equivalence of alternate survey formats (paper- and computer- based) of the measure.

Results of study 1 demonstrates that the M&MS satisfactorily discriminates between a community or low-risk sample and a clinic sample. The sensitivity of the measure was found equivalent to results found for the SDQ which is a widely used measure of community mental health for ages 11 and above. However, results suggest a reconsideration of the clinical thresholds of the emotional difficulties scale to allow better screening capacity, especially when the paper version of the measure is used, which is suggested by the results of the second study that found that scores from the paper survey were generally lower than those for the computer based survey. The results of the two studies, when considered together, support the need for reconsidering the thresholds and suggest that in future format specific thresholds might need to be developed for the paper version of the survey.

The second study investigated the equivalence of the computer and paper based survey formats of the measure. The measure had been developed and validated first as a computer based survey and the properties of the paper based measure could not be assumed to be the

same without prior testing. The results of this investigation indicate that the computer and paper survey formats result in different scores, with lower overall scores when using the paper survey. Additionally, this study utilised both scale- and item-level analyses to examine format equivalence and found that for this measure differences were mainly at the scale level. The results of the study have implications for other self-report mental health measures and the comparability of scores from different formats especially when non-randomly allocated in trials and in longitudinal studies. The findings also suggest the need for norms and thresholds that are survey format specific. This issue requires further investigation in future research both with this and other measures, especially as computer based testing and data collection becomes more common in the future.

Combined with the existing validation of the M&MS questionnaire (Deighton et al., 2013), these two studies further the psychometric examination of the M&MS measure and broadly support the utility of the measure as a screening tool in a school setting. However results also suggest the paper version of the measure might need to be validated separately especially in terms of establishing cut-offs. For the following studies in this thesis the data used had all been collected using the computer based version of the survey and hence existing thresholds that were established for the computer survey will be used.

Part 2: Externalising and internalising symptoms: developmental trajectories, co-development and underlying structure

Background

Developmental psychopathology

Developmental psychopathology is defined in varied ways. Lewis (2000) defines it as "*the study and prediction of maladaptive behaviours and processes over time (pp.3)"*, whereas it was defined by Sroufe and Rutter (1984) as "*the study of the origins and course of individual patterns of behavioural maladaptation, whatever the age of onset, whatever the causes, whatever the transformations in behavioural manifestation, and however complex the course of the developmental pattern may be*" (Sroufe & Rutter, 1984, pp.18). The various definitions have in common the acknowledgement that the field involves the study of development and psychopathology. Cicchetti and Toth (2009) describe it more in terms of a framework that allows for better examining links and drawing out theories in the study of psychopathology and development. It is an integrative field with links to varied disciplines including epidemiology, genetics, neuroscience, psychiatry, experimental psychology, sociology etc. alongside the study of mental health and developmental psychology (Cicchetti & Toth, 2009).

Child psychopathology is studied in two main contexts: the clinical context, usually within specialist mental health services, and the community context. It has been noted by many studying developmental psychopathology that although the clinic setting is interesting and provides detailed information about disorders, treatment, development, the community setting is vital as psychopathology and its development can only be understood in the context of the wider population, developmental norms and hence deviation for the norms (Cicchetti, 1984). From an epidemiological perspective, only community based studies allow us to estimate prevalence of disorder in the population. Additionally, they allow an examination of patterns of disorder in the community, which allow us to generate hypothesis about the causes and risk factors of disorders (Costello, 2009). A few decades ago, in the study of child mental

health we primarily studied cross-sectional samples, which provided useful clues regarding the aetiology and risk factors associated with developing mental disorder.

Understanding causal relationships is one of the key aims of longitudinal research, and as noted by Pearl (2009), forms one of the fundamental building blocks to understanding the world. Recently, the availability of longitudinal data sets has allowed us to use population based epidemiological data more effectively to try and extend our knowledge and hypotheses to the causal relationships between factors (Costello, 2011). Community based longitudinal studies have many strengths including: 1) greater generalisability of findings compared to studies focussing solely on clinical populations, 2) examination of effects of environmental including societal factors on the development of disorder, 3) the potential to observe those who do and do not develop disorder permitting better understanding of resilience and possible protective factors, 4) the study of individuals with sub-clinical levels of symptoms, and those who do not receive specialist treatment although they meet clinical criteria (referral biases), and 5) as highlighted before, allow the estimation of causal factors and cause-effect relationships in the development of psychopathology. Hence, in the study of the development of psychopathology, information regarding the continuities and discontinuities of symptoms provide crucial information to understanding the aetiology and mechanisms that play a role in the development of psychopathology.

Maughan and Rutter, (2011) differentiate three main topics from a methodological perspective that require longitudinal data. First, determining the factors associated with variations in development and outcome of childhood disorders. Second, the relationships between different types of disorder over time and the possible risk created by one form of psychopathology for another. Third, studying the risk factors and experiences in childhood and the adult outcomes of both those who suffer from the sequale of these risks and those who don't.

The focus in the following studies is going to be on the first two topics. First, identifying variations in development of internalising and externalising symptoms and the factors associated with them. Second, examining the longitudinal relationships between these two domains during development.

The study of variations of differential development of symptoms is crucial to not only understanding the development of symptoms but also understanding risk and resilience and factors that contribute to risk and resilience in different individuals (Rutter, Champion, Quinton, Maughan, & Pickles, 1995). In the study of the differential development of psychopathology the use of person-oriented approaches, i.e. approaches where the individual is the focus rather than a variable, are receiving greater attention recently as they can lead to greater understanding of developmental mechanisms by allowing a shift in the focus from the variables of interest to patterns of individual development (Bergman, von Eye, & Magnusson, 2006). The approaches based on this framework include pattern-based methods of analysis such as identifying heterogeneous trajectories of growth or estimating individual developmental slopes or that help identify different groups of individuals with different developmental pathways and then to examine the different predictors and risk factors of these groupings (Bergman et al., 2006).

Complexity in symptom development

Conventional approaches based on the general linear model, have used a 'single-trajectory' approach to modelling change which rests on the assumption that as individuals come from a single population a single growth trajectory adequately captures change (Jung & Wickrama, 2008). Although in overall growth trajectory approaches individual slopes are allowed to vary from the overall trajectory, this variation is difficult to capture in assessing the correlates of the slope and the impact of the slopes, where homogeneity is still assumed.

Hence, the homogeneity assumption of these techniques encompasses the assumption that covariates that affect growth influence each individual the same way.

In the absence of a single population trend for change in mental health symptoms across childhood and adolescence existing single-trajectory estimates make it difficult to understand how subgroups of individuals change as it can obscure distinct differences in how subgroups have different trajectories of change (Sterba et al., 2007; von Eye & Bergman, 2003). The possible pitfalls of assuming a single homogenous trajectory is illustrated by mixed findings of the single trajectory approaches of internalising symptoms in the same age range (childhood to pre-adolescence), with some studies finding increased symptoms over these years (Gazelle & Ladd, 2003), some stable (Keiley et al., 2003) and some decreasing (et al.Green et al., 2005, Wolpert et al., 2011). In reality the likelihood is that the average trajectory includes some decreasing, increasing and stable trajectories and the sample mean reflects the largest grouping. Studies in other domains of development, especially those focusing on behavioural problems, have been using a multi-trajectory approach to understand problem behaviours and this accounting for heterogeneous growth trajectories has improved the understanding of the different patterns and risk factors of behavioural problems in CYP (e.g., Roismann et al., 2010).

Different individuals experience varying symptom development; studying these individual trajectories is important as it not only increases understanding of the aetiology and development of psychopathology (Sroufe & Rutter, 1984), but also contributes to strategies of screening and treatment planning (Rushton, Forcier, & Schectman, 2002). Studies that have identified heterogeneous trajectories capturing individual differences in symptom change conclusively show not only that more clinically meaningful groupings can be achieved, but also that the heterogeneous approaches better represent the data in both externalising (Broidy

et al., 2003; Moffitt & Caspi, 2001) and internalising symptomology (Côté et al., 2009; Sterba, Prinstein, & Cox, 2007).

Correlates

An important aspect in studying epidemiological patterns and mechanisms has been identifying the various factors, including socio-demographic, that influence the risk of developing disorder (Verhulst & Koot, 1992). Examining the predictors of development of symptoms over time offers invaluable clues about different individuals' differential risk of developing, maintaining and recovering from symptoms.

Studies in the past tended to focus on the predictors of the whole-population trajectory. These approaches mask individual variations in change and assumes a common relationship between symptoms, time and predictors for all individuals in a population; this gives results that can be atypical of any individuals in a population, and hence, are not very useful (von Eye, 2010). The alternative, which has gained popularity in the last decade is the study of correlates of heterogeneous trajectories of symptom development and change. This has led to additional insights regarding the risk factors associated with symptom development, especially in regards to externalising behaviours (e.g., Odgers et al., 2008).

The current studies will include gender along with other socio-demographic and educational predictors such as ethnicity, deprivation, relative age, special educational needs and educational attainment. These variables are included in analysis as they are established predictors of differences in mental health symptoms and/or academic attainment. The specific associations with internalising and externalising symptoms are outlined in the backgrounds to the respective studies.

Cross domain relationships in development

In the past development theory first focussed on nature vs. nurture subsequently leading to focus on both genetic and environmental factors and their interactions as being

relevant to explaining development and risk of psychopathology (Lerner, Agans, DeSouza & Gasca, 2013; Lerner, 2006). Sroufe (2007) eruditely notes that genes and environment aside, it is the individual's cumulative developmental history that determines how future development will unfold. Increasingly, concepts and theories drawn on 'developmental systems theories' (e.g., Thelen & Smith, 2006) which depict developmental processes as involving relationships, which are bidirectional, among variables from the different levels that comprise the ecology of human development (Bronfenbrenner, 2006). This approach stresses on development being a function of interdependence and co-development between different domains and multiple levels of influence. Hence, interest in the links between key domains of development, such as mental health and academic attainment, is theoretically based on the assumption that different domains are linked developmentally (Thelen & Smith, 2006). Positive development or success in one domain is expected to provide scaffolding for positive development in the same and other domains; and conversely, negative development or deficits in one domain can result in negative development in the same and other domains (Masten, Burt, & Coatsworth, 2006).

Academic attainment predicts several varied lifelong outcomes including employment, health, life expectancy and human capital (Adams, 2002; Barro & Lee, 2001; Gregg & Machin, 2000; Guralnik, Land, Blazer, Fillenbaum, & Branch, 1993). Due to the life-long impacts of development in these two domains during childhood and adolescence, there is great interest in understanding the associations between psychopathology and academic outcomes over time. The cross sectional links between these two domains are well-established, with studies consistently indicating a negative association. However, the longitudinal associations have proven much more difficult to tease out. Studies have used a variety of different approaches to study the links between academic outcomes and symptoms of mental disorders over time (outlined in detail in the introduction to studies 3 and 4). However, the longitudinal links still remain unclear, especially for internalising symptoms,

with studies demonstrating either no relationship or negative associations over time (Riglin, Frederickson, Shelton, & Rice, 2013; Romano, Babchishin, Pagani, & Kohen, 2010). The inconclusive results are not necessarily due to different geographical areas, different age groups or different measures as inconsistent results have been found when these factors are similar (Romano et al., 2010). The lack of consistent findings and clear understanding of these relationships over time also have implications for early interventions in schools, resource use and intervention planning in schools, especially in light of recent relative increases in the stress on academic outcomes with other outcomes taking a back seat (Greenberg, 2010; Shoshani & Steinmetz, 2013).The current studies explore the possibility of using person-oriented heterogeneous trajectories to increase understanding of the associations and impact of symptom development on attainment in late childhood and early adolescence.

Co-morbidity and co-development of externalising and internalising symptoms

Externalising and internalising symptoms are the two broad domains that have been successfully employed in the study of child psychopathology. The reason for their endurance in the field is because they represent broader domains that include several specific diagnoses, which in children are not as highly differentiated as in adults. However, these domains that are treated distinctly in most literature in child psychopathology are moderately associated. The evidence for their associations over time in terms of development are sparse, compared to cross sectional studies of symptom/domain overlap. Many studies have focused on the longitudinal relationship in terms of which symptoms are pre-cursors to which (Cerdá, Sagdeo, & Galea, 2008), or in some cases the cyclical nature of the symptom development (Masten et al., 2005). Although there is insufficient consistency to be conclusive, studies suggest that to a greater extent externalising symptoms precede internalising symptoms (Cerdá et al., 2008), which considering externalising is to a greater degree childhood-onset and internalising adolescence-onset may not be surprising (Martel, 2013). Although there has been increasing focus on understanding longitudinal 'which-comes-first' relationships

between these domains with the availability of longitudinal data, there is relatively little study into the co-development of these symptom domains over time. Hence, the associations in symptom developments in these two domains is relatively unknown, especially in community-based samples. Hence, the study of associations in symptom concurrence, development and prediction across internalising and externalising domains in late childhood and early adolescence will be the focus of one of the following studies.

The existence concurrently or sequentially of co-morbidity is one of the biggest challenges to the diagnostic system in child psychopathology, in both research and practice (Jensen, Hoagwood, & Zitner, 2006). This issue is reflected in the research focussing on co-morbidity of symptoms (Caron & Rutter, 1991; Jensen, 2003), where-in there is greater awareness of the high extent of co-morbidity of symptoms in young people and the challenges this posits to the diagnosis based classification systems that are widely accepted, such as the DSM and ICD, whose use is increasingly championed as the standard best practice. These diagnostic classifications, mainly developed to increase reliability in diagnoses, some argue are absolutely necessary to then be able to develop empirically supported models of aetiology and development of symptoms (Jensen et al., 2006). The flip side of this argument is the view that excessive emphasis on these diagnostic classifications risk resulting in narrow, theoretically impoverished models of aetiology and development of mental disorder. This issue is also reflected in one of the key conceptually based discussions in psychotherapy-categories vs. dimensions (Zachar & Kendler, 2007), whether the underlying construct can be distinctively or categorically different from normative behaviour versus the view that disordered behaviours lie on a continuum and diagnosis reflects the end of this continuous dimension. In the absence of evidence to support strict dichotomies, some researchers argue for a dimensional view of psychopathology. However, the use of diagnostic classification systems such as the DSM- which force dichotomy (diagnosis or not) stands in direct contrast to the dimensional approach to psychopathology. Alternatively, some propose the

categorisations can be based on cause, rather than description of symptoms, an idea based on medical thinking (Zachar & Kendler, 2007) which as our knowledge stands today is not a feasible model. Psychopathology is characterised by situations where the same causes result in different diagnosis (multi-finality) and also situations where different causal or risk factors lead to same symptom manifestations (equi-finality)(Cicchetti & Rogosch, 1996). Research also indicates that the majority of known risk factors of psychopathology are common to most disorders i.e. risk factors do not differentiate between disorders (Caspi et al., 2013; Werry, Reeves, & Elkind, 1987). This is further compounded by the recognition that individuals present with differential susceptibilities to environmental risk factors (Belsky, Bakermans-Kranenburg, & van Ijzendoorn, 2007; Rutter et al., 1995), where certain risk factors or environmental factors do not affect all individuals in the same way, with individual differences in whether a risk factor leads to disorder or not.

The themes of dimension and diagnosis are key conceptual issues in the study of development and psychopathology, which have the possibility of being resolved or united within a new paradigm. The development of models that focus on processes instead of behaviour and symptom description has been suggested (Jensen et al., 2006), alongside suggestions that focus on improving the nosologies we already have (Zachar & Kendler, 2007). Recently, studies in adults have identified a broader dimension of general psychopathology (Caspi et al., 2013; Lahey et al., 2012), that offer a new paradigm within which to understand both risk of psychopathology and the extent to which disorders are different and have distinct predictors. This has not yet been explored in younger samples of children or adolescents, which is the aim of the last study in this thesis.

Methodological advances and developmental psychopathology

Methodologies for studying development or change over time have grown substantially and are increasingly accessible in the last few decades. This has been accelerated

by the importance placed on longitudinal data and cohort studies for understanding better the development and causal associations during development. These methodological advances in turn have vastly increased our understanding of the development and correlates of psychopathology. A case in point is Moffitt's theory of adolescence onset and childhood limited externalising symptoms (1993), which when proposed could not be explicitly tested. Inspired in part by the desire to test this hypothesis, Nagin and colleagues developed latent class growth analysis technique to identify individuals with different developmental trajectories (Bauer, 2007). Since, the method has been used on many different longitudinal datasets and increasingly consensus is being reached on the distinct types of trajectories, proportions of individuals, associated characteristics and their associated outcomes. This methodology, which is utilised in some of the studies that follow in this thesis, permits the identification of individuals with different distinct patterns of longitudinal development and has led to uncovering more information about longitudinal risk and its mechanisms and impact and hence increased our understanding of developmental complexities.

Hence, increasing theoretical knowledge in developmental psychopathology is coupled with advances in methodological and statistical approaches which can allow for the testing of theories or even help examine data in new ways leading to theory development. The following studies, where appropriate and necessary, utilise newer approaches to investigating the research questions with the hope that they might help in clarifying existing knowledge or uncovering new information to further our understanding of the development and structure of child psychopathology.

The current studies

The current studies first aim to explore complexity in both externalising and internalising symptom development followed by co-development of symptoms in these two

domains. This is followed by a re-examination of the underlying structure of child psychopathology.

Keeping with these aims some of the research questions that I aim to answer over the course of this work are:

- What are the observable variations of short term internalising and externalising symptom development in late childhood and early adolescence?
- Do the different trajectories observed have distinct correlates and differential impact on academic attainment?
- What is the relationship between externalising and internalising symptom development?
- Is there an underlying factor of psychopathology that explains the moderate associations between the internalising and externalising domains?

Study 3 examines the variations in development of *externalising symptoms* from ages 8-11 years and 11-14 years in two cohorts. The socio-demographic correlates of variations in symptom development are explored followed by an assessment of their relative impact on academic attainment over the same time period. Study 4, similar to the third study, examines variation in development of *internalising symptoms;* followed by similarly investigating the correlates and impact on academic attainment.

Study 5 focuses on the co-development of internalising and externalising symptoms in both the age groups over three years, both examining patterns in variations of symptom development across the two domains and also examining the associations between baseline scores and symptom development across the two domains.

The final study in this thesis, Study 6, based on the results of the previous studies, explores the possibility of a general psychopathology risk factor as an explanatory factor in both understanding co-morbidity in symptoms both cross-sectionally and over time. The study investigates the utility of this general psychopathology risk factor, a higher-level dimension,

by examining the correlates of a model where the general factor is included alongside the more specific internalising and externalising domains. The predictive capacity of the general psychopathology factor on future functioning, both psychopathology and academic, is also assessed.

Study 3. Short-term externalising symptom trajectories: correlates and differential impact on academic attainment

Background

Externalising disorders refer to disorders which are characterised by dysregulated behaviour (Kovacs & Devlin, 1998) and include symptoms of conduct disorder, oppositional defiant disorder, anti-social behaviour and substance mis-use (Achenbach, 1991; Goldberg & Goodyer, 2005). Alongside internalising symptoms they constitute the key broad domains of child and adolescent psychopathology (Kovacs & Devlin, 1998). According to population estimates externalising disorders are less prevalent than internalising symptoms (Green et al., 2005; Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993), and range from 5-12% of the population, varying depending on age (ibid.). However, due to the disruptive nature of these behaviours to family, society, school and individual environments, externalising symptoms are estimated to account for larger proportions of referrals to specialist mental health treatment (Dishion & Patterson, 2006; Kazdin, 1993).

Longitudinal development of externalising symptoms

In the BCAMHS cohort in the UK the rates of conduct disorder from late childhood to early adolescence almost doubled (Ford et al., 2003), whereas other studies have found similar or decreasing rates (e.g., GSMS, (Costello, Mustillo, et al., 2003). Substance use rates also significantly increase from childhood to adolescence (Costello, Copeland, & Angold, 2011). On the other hand rates of externalising disorders showed marked decreases moving from adolescence into adulthood (Copeland, Shanahan, Costello, & Angold, 2009; Costello et al., 2011), except substance use which increases into young adulthood (Copeland et al., 2009).

The study of the development of externalising symptom trajectories during childhood and adolescence is prolific, with many studies (compared to internalising) exploring symptom development, risk factors, impact and theoretical models explaining heterogeneous development of symptoms. The prevailing theories of developmental trajectories of externalising behaviours through childhood and adolescence are based on the taxonomy

proposed by Moffitt (1993), who first made the distinction between persistent externalising behaviours through the life-course and externalising behaviours that occurred only during adolescence (adolescence-limited). She posits that adolescent-limited externalising behaviours are not maladaptive, but rather normative and adaptive of adolescence and socialisation during these years. Person –centred trajectory approaches have largely been used to develop and validate this taxonomy of the development of anti-social behaviour.

This theory postulates that there are distinct types of externalising symptomology based on age of onset of symptom development, period of persistence and through to when it persists. Although the exact number of types of trajectories varies between three and five in the various studies (Moffitt & Caspi, 2001; Odgers et al., 2007; Roisman, Monahan, Campbell, Steinberg, & Cauffman, 2010), four trajectories have been found and are sufficiently established in the literature. Low symptoms, early onset and persistent, childhood limited and adolescent onset/limited (e.g., Barker & Maughan, 2009; Odgers et al., 2007; Roisman et al., 2010). The first group, low symptoms, is characterised by young people who demonstrate no or low levels of externalising symptoms through childhood and adolescence and usually consist of between 40% and 65% of the sample. Early onset and persistent is the group that has externalising behaviour throughout and this is associated with many negative consequences including pathological personality. Studies of long-term trajectories variously identify 5-11% of the sample that belong to this category (Barker & Maughan, 2009; Odgers et al., 2007; Roisman et al., 2010). The adolescent onset-limited symptoms group consist of 10 - 20% of adolescents and is used to refer to the group that develops symptoms only for a phase during the maturity gap phase of adolescence when physical maturity and social responsibility are incongruent (Moffitt, 1993). It is hypothesised that during these years externalising behaviours might prove adaptive but towards the end of adolescence these behaviours cease to be rewarding leading to a cessation of externalising behaviours in these individuals before adulthood. The last category was a later addition to the taxonomy as this

group was consistently found in data, the individuals with childhood-limited symptoms, which is characterised by symptoms during childhood that do not persist into adolescence. There are indications that although these individuals do not have life-course persistent symptoms, they still experience higher levels of life-time functioning impairment compared to the adolescent limited group (Raine et al., 2005).

This body of exploration has led to an increase in our understanding of the different risk factors, predictors and consequences of the identified trajectories of externalising or anti-social behaviour. However, existing research employing person centred trajectory based analyses, to a large extent has focussed on long-term trajectories and their predictors, typically studies have looked at trajectories over 12-40 years (Barker & Maughan, 2009; Odgers et al., 2008). The focus in these studies has been more on identification and examining correlates and risk factors of heterogeneous externalising trajectories and less on the predictive impact of different trajectories, and where it has looked at impact, the focus has been on long term adult consequences of childhood and adolescent externalising trajectories (e.g., (Odgers et al., 2008; Piquero, Farrington, Nagin, & Moffitt, 2010). For instance, Odgers et al. (2008) examined mental health, economic and anti-social behaviour outcomes and Odgers et al. (2007) examined physical health outcomes of the different trajectories at age 32 years and Piquero et al. (2010) examined impact on life failure reported at age 48 years. The use of shorter-term developmental trajectories to assess impact in the same time-period on other developmental domains remains an under-explored approach in this area of research.

Socio-demographic correlates

Gender. Gender differences in externalising behaviours are well-established with females exhibiting fewer externalising behaviours than males at all ages (Lahey et al., 2000). Studies have demonstrated higher prevalence in males (Maughan, Rowe, Messer, Goodman, & Meltzer, 2004), with some studies suggesting that boys demonstrate these problems 2-4 times more often as girls (Moffitt, Caspi, Rutter, & Silva, 2001). These gender differences in

externalising symptoms have been associated with gender differences in temperament from early stages of development (Muris & Ollendick, 2005; Ruble, Martin, & Berenbaum, 1998), with girls being more empathetic and pro-social (Zahn-Waxler, Robinson, & Emde, 1992). Evolutionary and biological explanations for gender difference in externalising also exist (e.g., sexual selection theory, (Martel, 2013)), with suggested links with puberty and hormones (Martel, 2013). Within females pubertal timing (at younger age) has been linked with higher risk for externalising and delinquent behaviours (Harden & Mendle, 2012).

In terms of the identified trajectories of development, some studies suggest that males are almost ten times more likely to have early onset-persistent symptoms (Moffitt & Caspi, 2001), contrarily some studies have found that the proportion in this group was almost equal (e.g., Barker & Maughan, 2009). Overall, males tend to represent higher proportions of all higher symptom trajectories. However, examinations of risk factors contributing to trajectory membership suggests that risk factors and associations between age of onset and development of symptoms are the same for male and female membership to certain trajectory groups (Barker & Maughan, 2009; Mazerolle, Brame, Paternoster, Piquero, & Dean, 2000; Moffitt et al., 2001).

SES. The association between socio-economic deprivation and externalising symptoms and behaviours is well-established in both cross-sectional studies (Dodge, Pettit, & Bates, 1994; Green et al., 2005; Patterson, Kupersmidt, & Vaden, 1990) and longitudinal studies (Costello, Compton, Keeler, & Angold, 2003; Macmillan, McMorris, & Kruttschnitt, 2004). Macmillan et al. (2004), established that poverty is intrinsically linked with externalising and anti-social behaviours with an investigation into changes in maternal poverty status and development of ext symptoms of children over four waves. They also found that the age or developmental stage in which children were exposed to more deprivation impacted on the amount and development of symptoms, for instance, children who were exposed to poverty around age 4 but then came out of poverty by age 7 years had

similar levels of problems as children who never experienced poverty (ibid.). Similarly, Costello, Compton, et al. (2003) using a natural experimental situation in the USA established that poverty is linked to externalising symptoms, which reduced in populations that were moved out of poverty (Costello, Compton, et al., 2003), with benefits of higher income that persisted into adulthood (Costello, Erkanli, Copeland, & Angold, 2010). One of the suggested mechanisms of this relationship is based on findings that parental monitoring increases when deprivation decreases, leading to more supervision of children which might account for some of the differences in externalising symptoms (Costello, Compton, et al., 2003; Jenkins, Rasbash, & O'Connor, 2003).

The associations between deprivation and the externalising symptom trajectories indicate that deprivation predicts greater risk of having higher symptom trajectories compared to having low symptoms over time; especially studies consistently find that deprivation is associated with greater risk of persistent symptoms through childhood and adolescence (Moffitt & Caspi, 2001; Odgers et al., 2007).

Ethnicity. Links between socio-ethnic grouping and behavioural problems have been explored to a greater extent and some consistent links have been found. Studies in Britain have found that antisocial behaviours are significantly higher in African –Caribbean Black and lower in Asian individuals (Goodman, Patel, & Leon, 2008; Smith, 2005). Social and racial discrimination and deprivation are discussed as some explanations, but as noted by Nikapota & Rutter, (2009) both are minority groups that face racial discrimination and social disadvantage in the UK but the former has higher rates and the latter lower. The underlying mechanisms for differences have started being explored and have not yet led to any clear explanations (e.g., Goodman, Patel, & Leon, 2010), although cultural differences are suspected to play a role (Nikapota & Rutter, 2009).

Relative Age. The effect of relative age, or age-within-cohort, is often studied within the school context on academic outcomes, however it is often left out from population based epidemiological studies (Costello, Mustillo, et al., 2003; Green et al., 2005), although they have been documented in other child developmental domains such as self-esteem (Thompson, Barnsley, & Battle, 2004) and academic performance (Boardman, 2006; Fredriksson & Öckert, 2005). The negative effect of relative age in relation to overall mental health difficulties was found in a large cross-sectional study of mental health (Goodman, Gledhill, & Ford, 2003). Lien, Tambs, Oppedal, Heyerdahl, and Bjertness (2005) studied relative age effects at the end of secondary school in Norway and found that being younger negatively impacted on peer relationships of boys and did not seem to have a negative impact on internalising and externalising symptoms. Effects of being younger within the cohort have neither been explored for sub-domains in childhood and early adolescence nor in the longitudinal context for mental health.

SEN. The relationships between special educational needs, considered as an aggregate and externalising symptoms is difficult to establish, especially considering that having disruptive externalising symptoms can lead to a special educational need statement within schools in England. Studies have found that young people with intellectual disability and autism spectrum disorders have high rates of externalising disorders (Emerson, 2003; Simonoff et al., 2008) and high associations are found between learning difficulties and externalising problems from early primary school years (Prior, Smart, Sanson, & Oberklaid, 1999). In terms of the developmental trajectories of externalising symptoms, studies that have focussed on the SEN correlates of symptom development were not identified.

Externalising symptoms and academic attainment

Interest in the links between externalising symptoms and educational attainment have been the focus of much research and is of great interest for reasons that were outlined over two decades ago, but are still pertinent today, by Hinshaw (1992) and can be summarised as

follows: 1. They constitute major problems of childhood both in terms of prevalence rates and personal and societal impact (Green et al., 2005), 2) both domains of development strongly predict later outcomes, health, social and economic (Adams, 2002; Crystal, Shea, & Krishnaswami, 1992; Odgers et al., 2007), 3) understanding underlying mechanisms can increase *"theoretical insights into behaviour-cognition links in both normal and atypical development"* (p.127, Hinshaw, 1992), and 4) the associations have direct implications for policy and practice in the fields of education and other child-related areas (Hinshaw, 1992). Better understanding these associations can contribute to the debate surrounding the increasing role of schools in providing whole child development including focussing on their well-being (Greenberg, 2010), alongside their more traditionally understood role of school as a place of learning. Especially considering the disruption caused by students with externalising symptoms to classroom environment and learning (Figlio, 2007) and evidence that school-based early intervention can help reduce externalising symptoms (Deighton, Patalay, et al., 2013), better understanding these links has the potential to contribute to policy and practice around school based support.

The inverse associations between externalising symptoms, and educational attainment has been consistently demonstrated in several studies (Ansary & Luthar, 2009; Hinshaw, 1992; Kessler, Foster, Saunders, & Stang, 1995; Masten et al., 2005; Moilanen, Shaw, & Maxwell, 2010). Studies also indicate that these two domains are linked from a very early age, even before schooling begins (Hinshaw & Anderson, 1996).

Studies examining this relationship longitudinally have largely found that externalising symptoms predict low subsequent educational attainment (Hinshaw, 1992; Moilanen et al., 2010) and this relationship is considerably more established in the literature compared to the one between internalising symptoms and attainment. However, Duncan et al. (2007) analysed six different datasets from different countries (US, Canada, UK) to assess the impact of externalising symptoms at kindergarten entry on subsequent attainment and found

that after controlling for prior attainment, the impact of externalising symptoms was not significant in 4 of the studies and was in 2 of the studies.

Cascade modelling, which explores simultaneously the associations between domains within and across time in a developmental manner, has been increasingly used in recent years to explore longitudinal links between externalising symptoms and academic outcomes. van Lier et al. (2012) explored the impact of externalising behaviours on attainment and peer relationships from age 6-8 years and found that early externalising behaviours predict worse attainment and higher peer victimisation. Moilanen et al. (2010) similarly examined this relationship in boys from middle childhood to early adolescence and found that higher levels of externalising difficulties predicted low academic competence. Chen, Huang, Chang, Wang, and Li (2010) also examined these relationships from age 8-12 years and found that externalising behaviours predicted later academic attainment, and not vice versa. Contrastingly, Vaillancourt, Brittain, McDougall, and Duku (2013) examined these relationships from age 10-14 years and found support for academic attainment predicting subsequent externalising behaviours.

With regards to externalising symptoms and academic attainment, results from cascade modelling have lent support to two competing theories of the relationships between symptoms and academic learning: academic incompetence, and adjustment erosion. The first, the *academic incompetence* theory, posits that problems with learning and academics have a knock on effect onto behaviours via mechanisms including interactions with deviant peers, frustration and disaffection and reduced engagement. The contrasting hypothesis, the *adjustment erosion* hypothesis predicts that difficulties lead to subsequent academic difficulties (Moilanen et al., 2010). Although there is evidence for both theoretical hypotheses regarding the longitudinal relationships between these two domains and one can conclude that both are relevant and effect outcomes in a cyclical manner, there seems to be more support from cascade modelling for the adjustment erosion hypothesis (e.g., Chen et al., 2010). The

current study will focus on the adjustment erosion hypothesis with the aim of better understanding the relationships between heterogeneous symptom development and attainment.

In summary, in spite of some inconsistency in findings, the negative impact of externalising symptoms on academic attainment is a fairly established finding, with more studies indicating negative longitudinal impact (egs. Chen et al., 2010; Hinshaw, 1992) than not (Duncan et al., 2007). However, the commonality in the existing research approaches is a lack of studies that have studied differential impact of externalising symptoms in individuals with heterogeneous symptom trajectories. Current studies have examined this impact aggregated over the whole sample which does not permit a more nuanced breakdown of different trajectories on impact. As outlined above, in studying the development of externalising symptoms there is a long tradition of studying trajectories of symptom development and their correlates. Extending this to studying their relative impact on attainment might help further understanding of the more immediate impacts of different developmental trajectories.

The current study aims to address these limitations and extend our knowledge of underlying mechanisms of links between these two domains by applying a person-centred followed by variable-centred methodology to examine the relationships between development of symptoms and later educational attainment.

The current study

Investigations of longitudinal trajectories of externalising behaviour have usually focussed on long-term trajectories and the risk factors associated with them (Odgers et al., 2008). The current study utilises a much shorter time-span and hence does not expect or aim to replicate existing known trajectories of externalising symptoms. Instead, the focus is on identifying heterogeneous patterns of development of externalising symptoms in a shorter

time span (3 time points, covering 2 years) as a way of summarising patterns in short term development of symptomology and their associated risk factors. Risk factors of interest include both socio-demographic factors and school-related factors thus making the research relevant to the school setting. The main interest in identifying these short-term trajectories lies in investigating the impact they have on another key domain of childhood development: educational attainment.

As outlined above, investigations into the longitudinal associations between externalising symptoms and educational attainment have utilised various different methodologies that have helped uncover different bits of the puzzle. A commonality in all these methods has been the 'lumping' or 'aggregating' of all individuals in a sample while carrying out analysis, the limitation of this being that it assumes identical relationships between time, risk factors and outcomes for all individuals in a sample. The current study takes a person-centred approach, which is represented by the trajectory studies, and uses the trajectories to predict relative change in attainment over the same time period. By breaking the sample down into different groups based on development of externalising symptoms we hope to examine the differential effects of different developmental symptom pathways on cross-domain outcomes.

It is hypothesised that heterogeneous externalising symptom development across three waves will impact differentially on change in educational attainment over the same period.

Method

Design

Externalising symptoms were assessed in the first term of school (which is during autumn in England) every consecutive year for three years. Educational attainment scores were taken from national standardised tests at the end of a Key Stage (KS) in England which correspond to age 7 (KS1), age 11(KS2) and age 14 (KS3) (see www.education.gov.uk/ for detailed descriptions of KSs). The manner in which the two educational attainment measurement points frame the three waves of data pertaining to participants' externalising symptoms provides a scenario where development of symptoms during the three waves can be used to examine the impact of development of symptoms on relative change (gains or losses) in national standardised tests of educational attainment.

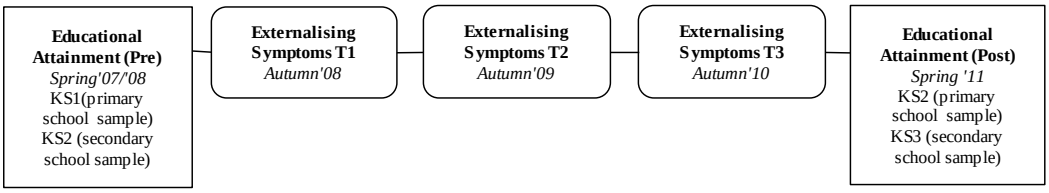


Figure 3.1. Design of the study illustrating when data for key variables, educational attainment and externalising symptoms, were collected

Participants

Data from a naturalistic three year longitudinal study of mental health in schools in England (Wolpert et al., 2011) were utilised in this study. Data were collected with yearly intervals from 138 primary and 37 secondary schools who participated in all three waves of the study. In wave one, 4961 primary and 5087 secondary school pupils participated in the survey. At the end of three waves complete data were available for 3346 pupils from primary and 2647 pupils from secondary schools, which are used in analysis. Attrition was mainly due to absenteeism and entire form groups in schools being unable to participate in some waves.

In the complete cases samples that are utilised in analysis (primary N=3346, secondary N=2647), at the first wave of data collection mean age of the primary school sample was 8.70 years (SD=.30). Almost half the sample was Female (49.3%) and free school meal (FSM) eligibility (18.3%) was higher than the nationally (14%; DFE 2010). In the secondary school sample mean age at first time point was 11.71 years (SD=.29). Slightly more than half the participants were female (54.4%) and free school FSM was higher than national levels, 17.1% (vs.11.9% nationally; DFE, 2010). In both samples the majority of participants were classified as White (primary: 76.6%, secondary: 73.6%) followed by Asian (primary: 12.6%, secondary: 17.5%), Black (primary: 4.7%, secondary: 4.9%), Mixed (primary: 3.7%, secondary: 3%) and other (primary: 2.5%, secondary: 1.1%). 8.6 % primary pupils and 7.1% of secondary pupils were classified as having special educational needs. Educational attainment scores in national standardised tests were similar to national levels with mean Key Stage 1 scores in the primary score sample (M=15.12, SD=3.47) slightly lower than the 15.3 national average, and in secondary school sample the mean Key Stage 2 score of 27.7 (SD=4.19) was identical to the 27.7 national average.

In terms of differences compared to the pupils lost to within-school attrition, in both primary and secondary school samples there were no differences in gender proportions (primary χ^2 =.04, p=.85; secondary χ^2 = .44, p=.51). The final samples had lower proportions of students eligible for free school meals (primary χ^2= 19.4, p<.001; secondary χ^2=7.89, p<.01) and with special educational needs (primary χ^2= 11.25, p<.001; secondary χ^2=17.53, p<.001) when compared to students who participated in wave one. Externalising symptoms at time 1 were significantly lower in the final sample in both the primary (t= 2.26, p<.05) and secondary school (t=3.56, p<.001) samples. In both primary and secondary school samples, educational attainment scores in national tests were higher in the final sample than in the wave 1 sample (primary t = 4.08, p<.001; secondary t= 3.79 , p<.001).

In summary, the final sample analysed in this study is representative of pupils nationally except for deprivation, as the study sample has a slightly higher proportion of deprived pupils. However, attrition indicates that pupils lost in follow-up waves were significantly more likely to be eligible for FSM, to have SEN, have higher externalising scores at baseline and to have lower attainment scores.

Procedure

Computer-based surveys were completed by pupils within the normal school day with support from their class teachers. Consent was sought from parents via mail beforehand each year using an opt-out approach. Class teachers facilitated online, whole-class survey completion sessions for children and were given a standardised instruction sheet to read aloud that outlined what the questionnaire was about, the confidentiality of their answers and their right to decline participation. The online survey system was designed to be easy to read and child-friendly. All participants received information about the study, including explanation of the confidentiality of their responses and their right to decline to participate and drop out at any time.

Measures

Externalising symptoms

Externalising symptoms were measured using the *Behavioural Difficulties* scale of the Me and My School questionnaire (Deighton, Tymms, et al., 2013), which is a 6-item self-report scale (e.g., 'I hit out when I'm angry') with three response options: never, sometimes, always. The answers to the items are summed to create a total behavioural difficulties score, higher scores indicating more difficulties. The scale has an at-risk cut-off score of 6 (Deighton, Tymms, et al., 2013).

Academic Attainment

National standardised test results, referred to as Key Stages (KS) in England, were used as a measure of attainment. Score averages of English, mathematics and science were used as a measure of attainment. For the primary school sample combined KS1 scores (mean=15.12; SD=3.47) were used as a measure of attainment prior to the three waves and the KS2 score (mean=27.76; SD=4.21) was used as measure of attainment post the three time points. For the secondary school sample, KS2 average level (mean =4.16, SD=.66) were used as a measure of attainment prior to the three waves and KS3 average level (mean=5.54, SD=.96) was used as measure of attainment post the three time points. 110 primary and 143 secondary students did not have either one or both attainment measures and are hence excluded from Stage 3 analysis. The scale is different for the two cohorts as KS1 scores are available only as points, KS2 scores are available as both levels and points and KS3 scores are available only as levels. Hence, to ensure a uniform scale of measurement within a sample average points were used in the primary schools sample and average levels in the secondary school sample.

Correlates

Socio-demographic information was derived from the National Pupil Database, which holds all school-related data pertaining to every student in England.

Gender. Participants reported on their gender and this information was cross-referenced with school help information and the NPD to create a gender variable. In all analysis males were coded '0' and females '1'.

Ethnicity. Ethnicity information was divided into the broad categories of White, Asian, Black, Mixed and other (other consisted of participants belonging to groups with very low proportion [e.g., Gypsy, 0.1%], refusing information, or their ethnicity code recorded as being unclassified).

Socio-economic status (SES). SES was measured by free school mean eligibility (FSM), which is a widely used proxy for deprivation in school-based research (Hobbs &

Vignoles, 2010). The binary variable was coded '0' for not FSM eligible and '1' for participants who were eligible for free school meals.

SEN. SEN status was recorded from the National Pupil Database (NPD) and students were included as having SEN if they had either a SEN statement or school provision to support SEN (in NPD referred to as Statemented and School Action Plus). Individuals were assigned '1' if they had either statement of school action plus and '0' if not.

Age. Age was estimated to the month from month and year of birth data available for each participant at wave 1 of data collection. As chronological age was accounted for by time points year on year, the age variable in the models represents age within the cohort at any time point, hence representing relative age within the cohort.

Analytic strategy

The main objective of modelling longitudinal developmental trajectories is it allows us to capture information about "inter-individual differences in intra-individual change over time" (Nesselroade, 1991). Conventional approaches have used a 'single-trajectory' approach to modelling change which rests on the assumption that as individuals come from a single population a single growth trajectory adequately captures change (Jung & Wickrama, 2008). The homogeneity assumption of these techniques also encompasses the assumption that covariates that affect growth influence each individual the same way. Contrary to this assumption, findings from several research papers in the social and health sciences looking at growth models over time in areas such as conduct problems (Roisman et al., 2010) and alcohol use (Greenbaum, Del Boca, Darkes, Wang, & Goldman, 2005; Jackson & Sher, 2005), indicate that there is a heterogeneity of growth trajectories and that using a single growth trajectory estimate 'over-simplifies' things. Methods like latent class analysis and growth mixture modelling aim to uncover the heterogeneity in a sample by finding substantively meaningful groups of individuals that have similar levels and trajectories of the

variable of interest (Muthén, 2004). To better understand the overall framework and differences between traditional regression methods and latent class methods an understanding of the different foci becomes relevant. Muthén and Muthén (2000) discuss the distinction between 'variable-centred' and 'person-centred' approaches. Methods based on the regression framework have a focus on describing the relationships between variables and identifying significant predictors of outcomes. On the other hand person-centred approaches like latent class analysis and cluster analysis classify individuals into groups or categories based on response patterns which results in grouping where individuals within groups are more similar to each other and different from individuals in a different group (Jung & Wickrama, 2008).

The current study aimed to use symptom development in the three waves between attainment measures as a predictor of change in academic attainment over that period of time. Recognizing that symptom development is heterogeneous in a sample, we decided to use empirically derived trajectories to summarize different developmental pathways over three waves. Hence analyses were conducted in multiple steps.

In the subsequent analysis a combined approach has been applied, whereby first a person-centred approach is taken to identify groupings or classes of similar individual trajectories over time. Subsequently a variable-centred approach is applied to the groupings to ascertain the relationships between variables and these classes to identify predictors of these classes. Analyses were done in three stages to 1) identify heterogeneous developmental trajectories; 2) investigate the correlates of different symptom trajectories, and 3) assess how trajectories predict change in academic attainment.

Stage 1: Identifying heterogeneous developmental trajectories

The two approaches available to identify developmental trajectories are *latent class growth analysis* (LCGA) and *growth mixture modelling* (GMM). Both approaches are person-centred and identify groupings of individuals based on common trajectories. They differ in

that LCGA identifies trajectories by both intercept and slope whereas GMM focuses on random slopes, and tends to identify groupings based on common slope (Muthén, 2006), which might render it better suited to more data points (the current study only has three waves of data). As a result of these slightly different approaches LCGA usually results in the selection of solutions with more classes when compared to GMM. It has been suggested that the decision is best made based on convergence and better model fit (Muthén, 2006). When trialling both approaches in the current data GMM models did not always converge and the neatness of classification, as measured by entropy, was lower (<.70). Given the aim in the present study was to use the person-centred approach to summarise development of symptoms over a small number of waves (3), only to be able to use them as predictors of other outcomes, the LCGA approach was used to identify trajectories. This had the advantage of providing a more heterogeneous higher class-solution with greater entropy which resulted in trajectory groupings that could be used as predictors in further analysis.

LCGA, a semi-parametric technique which identifies sub-groups of individuals following a similar pattern over time (Nagin, 1999), was conducted in Mplus7 (Muthén & Muthén, 2012) to estimate empirically derived trajectory models and identify a k-trajectory model that had good fit criteria, parsimony and theoretical interpretability. Criteria used to assess and select a k-trajectory model for further analysis included model fit, neatness of classification and interpretability (Jung & Wickrama, 2008). There are several criteria that are used to determine the number of latent classes that can be explored further (see Box 1). It is important to note that number of classes is not a finite, definite thing as there can be any number of classes between 1 to the maximum number of combinations possible with the measure of interest, but for the sake of interpretability a certain model with k classes might be chosen to allow for further exploration of the groupings or classes of individuals (Nagin & Tremblay, 2005).

Box 1: Overview of different criteria to determine number of latent classes

Likelihood ratios:
- Lo-mendell-rubin likelihood ration test (LMR-LRT)
- Bootstrap likelihood ratio test (BLRT)
- Chi-square difference test

Fit indices:
- AIC
- BIC
- Sample Adjusted BIC

Other criteria:
- Successful convergence
- High entropy
- High posterior probabilities

Most important
- Parsimony, interpretability and one's research question

As can be seen from the box above there are several likelihood ratio tests and fit indices that give information about the fit of the model. These indices compare a model with k classes with a k-1 class model to determine if the k model is significantly better that the k-1 class model. Fit indices are computed based on likelihood function and parameters in a model, with a lower value indicating better model. Fit indices, such as AIC and BIC, are extremely sensitive to sample size and hence might favour highly parameterised models (Marsh, Balla, & McDonald, 1988). With larger sample sizes the sample size adjusted BIC (A-BIC) is recommended has been found to be superior in simulated studies of information criteria (Yang, 2006). In interpreting BIC, a reduction of 5 or more is considered evidence for a large difference between two models (Raftery, 1995). A simulation study testing the different ratios and indices by Nylund, Asparouhov, and Muthén (2007) concluded that the BLRT was the most consistent indicator of classes and the BIC performed the best from the information criteria.

The other criteria that need to be considered to determine number of classes include successful model convergence, high entropy, 1% class membership and posterior probabilities (Jung & Wickrama, 2008). Of the other criteria, successful convergence is essential and is determined by testing model convergence with a greater number of starts, with the same best likelihood ratio indicating that model convergence has been achieved.

Entropy assesses whether individuals were neatly classified into one category or not (Wang & Bodner, 2007) and an entropy of 1 indicates that the whole sample neatly fell into one or another class and hence a score closer to 1 indicates neater classification. Posterior probabilities represent the likelihood of individuals being in a certain class (estimated class probability) and are computed by the average latent class probabilities for the most likely class membership (row) by latent class (column) and include values from zero to one with values closer to one indicating neater class classification, as it is from these probabilities that class membership is assigned (it is analogous to factor scores in factor analysis; (Clark & Muthén, 2009). As outlined above, and re-iterated by papers describing criteria for model selection; the aim of identifying trajectories is to find substantively meaningful groupings that are of theoretical interest to the researcher (Jung & Wickrama, 2008; Muthén, 2004)

Based on all these criteria and the research question and interpretability of the models, a k class model will be selected for further exploration. The selected model with k classes will then be presented with descriptives, intercepts, slopes etc. School level variation in class membership was also assessed using *glamm* in STATA (Rabe-Hesketh & Skrondal, 2012) to assess if further analysis needs account for nesting of pupils within schools or not.

Stage 2 Correlates of heterogeneous symptom trajectories

To explore the different predictors and risk factors associated with the heterogeneous trajectories identified multinomial logistic regressions (MLR) were conducted to assess how associations with correlates of trajectory-membership in comparison to the reference group. Based on popular convention, in the MLR the trajectory with the largest proportion of individuals was used as the reference group. For reasons outlined in the introduction gender, ethnicity, deprivation, relative age, SEN and academic attainment were included as predictors in the analysis. MLR were conducted in STATA12 (StataCorp, 2011)**.** Relative risk ratios (RR) that represent the probability for the predictor of interest of having a certain trajectory when compared to the reference group were estimated in STATA (using the *rrr* command

after conducting the MLR) to allow for easy interpretation (McNutt, Wu, Xue, & Hafner, 2003). A RR greater than 1 indicates that the risk is increased for the predictor category/unit change in predictor in question and, inversely, RR's less than 1 indicating reduced risk.

Stage 3 Predicting change in academic attainment

The third stage of analysis aimed to test the predictive ability of the derived trajectories to examine the relationship between different trajectories and change in academic attainment over the three waves. This was examined by looking at how trajectory groups predict attainment after the last wave after controlling for attainment prior to the initial wave. As school-level variation in attainment was high (>20%), a multi-level modelling (MLM) approach was used to account for nesting. MLMs were computed using both aggregated symptom scores and the derived trajectories as predictors to allow comparison of the predictive utility of the trajectory based approach. MLMs were computed in STATA12 (StataCorp, 2011) using the derived trajectories as categorical predictors by creating a dummy variable where, like in the previous stage, the group with the highest proportion was the reference group. Effect sizes for main effects were computed by dividing the beta estimate for main effect by the average of the square root of the variance estimate at both time points (Fonagy et al., 2009). The following models are presented:

1. Baseline model: A baseline model with all socio-demographic predictors was run as preliminary to further analysis
2. Aggregate symptoms: Model 1 with aggregated symptoms over 3 years to assess the extent to which aggregated symptoms across all individuals predict change in academic attainment
3. Trajectories: This analysis includes the baseline model along with the trajectories, coded as categorical variables, predicting attainment. The co-efficients indicate the effect of being in each trajectory group when compared to the reference category.

Results

All stages of the results are first presented for the younger (8-11 years) and then for the older sample (11-14 years). For ease of reading the Tables are presented in blocks at the end of each set of analysis.

Results: 8-11 years

Descriptive statistics

The means, standard deviations and correlations of all the study variables are presented in Table 3.1.

Stage 1: Identifying heterogeneous developmental trajectories

Model criteria and results from the LGCA are presented below in Table 3.2 followed by a rationale for the selected model for further exploration. Table 3.2 presents fit indices and criteria for model selection for the 2-7 class solutions obtained.

The 6-trajectory model was selected based on several criteria including the having the best neatness of classification (entropy=.8), log-likelihood differences indicated it was significantly better than the 5-class model and that the 7-class model is not a significant improvement on it. The 6-class model also showed sufficient heterogeneity with the largest class consisting of less than 50% of the population. The 6-trajectory model is presented in Figure 3.2 and Table 3.3 presents sample descriptives and trajectory intercepts and slope co-efficients for each of the 6-trajectories.

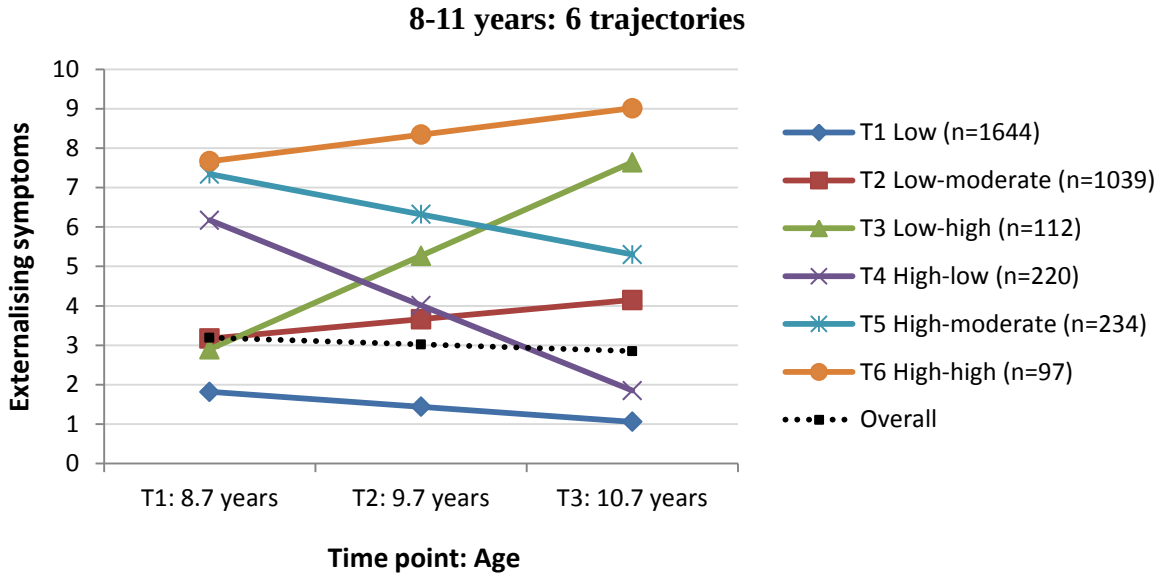


Figure 3.2 Heterogeneous developmental trajectories of externalising symptoms in children aged 8-11 years (6-trajectory model)

As can be seen from Table 3.3, proportions of children with different trajectories over the 3 waves varied greatly with the largest proportions of participants having low-low (49.1%) and low-moderate increasing (31.1%) symptoms over the 3 waves. The lowest proportions of children were identified as having increasing low-high symptoms (3.3%, N=112) and high-high symptoms (2.9%, N=97). 6.6% (N=220) and 7% (N=234) had decreasing high to low and high to moderate symptoms over the three years.

School level variation in trajectories was estimated to assess the need to account for pupils nested within schools in further analysis. The amount of school level variation in trajectories was small ICC= .04, and hence nesting was not accounted for in Stage 2 analysis.

Stage 2: Correlates of trajectory membership

MLR analysis was conducted with the trajectory that represent individuals with low scores at all three time points being used as the reference category as it was the trajectory group with the largest proportions. Table 3.4 presents the results of the MLR in the younger sample. Relative risks (RR) and their 95% confidence intervals are reported.

Females were significantly less likely to be in all the trajectory groups when compared to the reference low-low trajectory. Ethnic groupings by and large did not significantly predict higher risk of belonging to trajectories with the exception of Black ethnicity predicting higher probability of belonging to the low-medium increasing trajectory. Deprivation significantly predicted higher probability of having low-medium, high-medium and high-high trajectories with children with FSM being more than two and a half times more likely of having a high-high trajectory. SEN and relative age did not significantly predict group membership to any of the trajectory groups whereas educational attainment significantly predicted membership to all groups with higher attainment predicting lower probability of belonging to any of the other trajectory groups when compared to the reference group.

Stage 3: Predicting academic attainment

Table 3.5 presents the results of the MLMs in the younger sample, predicting subsequent attainment while controlling for prior attainment. The regression co-efficient and its standard error are both presented along with the effect sizes for the trajectories. As can be seen from Table 3.5, the baseline model (Model 1) indicates that when prior attainment is controlled for, individuals who are female, deprived, younger within their class and have special educational needs have significantly lower attainment scores and individuals classified as belonging to Asian or other ethnic categories have significantly higher later attainment.

Model 2 demonstrates the results when an aggregated symptom score is used to predict change in attainment. The negative significant co-efficient indicates that even when taken as an aggregate higher symptoms predict negative change in attainment (β = -.13, ES= .03). The results of Model 3 indicate that belonging to the different trajectory groups had a significant negative impact on subsequent academic attainment scores when compared to the reference low-low trajectory. In terms of magnitude of effects, the extent to which the trajectories impacted on attainment varied with the highest magnitude effect found for the low-high trajectory (β = -1.29, ES =.32). The low-moderate, high-moderate and high-high groups had similar co-efficients (β =.40 - .45, ESs ~10) and the high moderate trajectory predicted -.63 decrease in attainment compared to the low reference group (ES=.17).

Table 3.1. Means, standard deviations and correlations for variables in the primary school sample

	Mean (SD)/ Percentage	**Correlations**							
		1.	2.	3.	4.	5.	6.	7.	8.
1.T1 Externalising	3.19 (2.54)								
2.T2 Externalising	3.00 (2.47)	.55**							
3.T3 Externalising	2.85 (2.43)	.48**	.60**						
4.Pre-attainment	15.13(3.47)	-.19**	-.20**	-.18**					
5.Post-attainment	27.76(4.21)	-.18**	-.19**	-.17**	.72**				
6.Age at T1	8.70 (.30)	-.02	-.02	-.02	.19**	.09**			
7.Gender	49.3%	*-.22***	*-.22***	*-.25***	*.09***	*.04**	*-.04**		
8.FSM	18.1%	*.07***	*.11***	*.11***	*-.29***	*-.23***	*.010*	*.001*	
9.SEN	8.6%	*.13***	*.14***	*.13***	*-.34***	*-.31***	*-.06***	*-.14***	*.13***

**Italicised are non-parametric correlations (Spearman's rho)*

Table 3.2. Fit indices and model selection criteria for the 2-7 class solutions in the primary sample (8-11 years)

	% in classes	**Posterior probabilities**	**Entropy**	**A-BIC**	**BLRT**	**LMR-LRT**
2-Class	13 & 87	.84-.95	.76	43655.44	313.11***	300.76***
3-Class	9-78	.78-.92	.76	43441.35	228.90***	219.87**
4-Class	9-61	.77-.91	.74	43268.07	188.10***	180.68**
5-Class	3-54	.77-.90	.78	43161.220	121.661***	116.86
6-Class	3-49	.74-.91	.80	43044.27	131.77***	126.57***
7-Class	2.5-44	.71-.90	.77	42999.18	59.90 ***	57.53

<.10, **<.01, *<.001*

Table 3.3. Sample breakdown, socio-demographic descriptives and intercept and slope co-efficients for the trajectory groups in the primary school sample (8-11 years)

Trajectory group	**N (%)**	**Gender % (Female)**	**FSM % (Yes)**	**Age M (SD)**	**SEN % (Yes)**	**Attainment M (SD)**	**Intercept**	**Slope (p)**
T1 *Low-low*	1644 (49.1)	61.9%	14.5%	8.71 (.30)	5.3%	15.71 (3.24)	1.82	-.38***
T2 *Low-moderate*	1039 (31.1)	41.7%	19.9%	8.71 (.30)	9.6%	14.95 (3.53)	3.17	.49***
T3 *Low-high*	112 (3.3)	28.6%	20.7%	8.68 (.29)	15.3%	14.23 (3.44)	2.90	2.37***
T4 *High-low*	220 (6.6)	40.9%	21.0%	8.69 (.29)	13.2%	14.32 (3.67)	6.17	-2.16***
T5 *High-moderate*	234 (7)	25.2%	27.2%	8.71 (.30)	16.2%	13.70 (3.59)	7.34	-1.02***
T6 *High-high*	97 (2.9)	19.6%	36.5%	8.68 (.31)	21.9%	13.16 (3.37)	7.67	.67*
Overall sample	3346	49.3	18.3	8.70 (.30)	8.7	15.12 (3.47)	3.19	-.17***

<.10, **<.01, *<.001*

Table 3.4. Relative Risk Ratios (RR) for the Multinomial logistic regression of the 6-trajectory model in primary school sample

		T1 *Low*	**T2 *Low-medium***		**T3 *Low-high***			**T4 *High-low***			**T5 *High-medium***			**T6 *High-high***		
		RR (SE)	95% CI		RR(SE)	95% CI		RR (SE)	95% CI		RR (SE)	95% CI		RR(SE)	**95% CI**	
Gender (Female)		.45*** (.04)	.38	.53	.24*** (.05)	.15	.37	.45*** (.07)	.34	.61	.22*** (.04)	.16	.30	.16***(.04)	0.10	0.28
Ethnicity-Asian		1.14 (.14)	0.89	1.46	1.34 (.38)	.77	2.33	1.43 (.30)	0.95	2.15	.82 (.20)	0.51	1.31	1.13 (.36)	0.61	22.11
Ethnicity-Black		1.72** (.33)	1.18	2.51	1.09 (.59)	.38	3.12	1.98^(.62)	1.07	3.66	1.20 (.43)	0.59	2.44	.56 (.41)	0.13	22.37
Ethnicity-Mixed	**Reference group**	.94 (.21)	0.61	1.47	1.34 (.65)	.52	3.49	1.15* (.45)	0.53	2.47	.99 (.39)	0.46	2.15	.87 (.54)	0.26	22.93
Ethnicity-Other		.80 (.27)	0.41	1.55	.00 (.00)	.00	-	.83 (.52)	0.25	2.81	.00 (.00)	0.00	-	1.06 (.81)	0.24	44.73
Relative age		1.11 (.16)	.84	1.46	.86 (.30)	.44	1.69	.95 (.24)	.58	1.57	1.16 (.30)	.70	1.92	.91 (.34)	0.44	11.89
FSM(Yes)		1.35** (.15)	1.08	1.68	1.36 (.35)	.82	2.27	1.29 (.25)	.89	1.87	1.80*** (.32)	1.27	2.56	2.58***(.62)	11.61	44.15
SEN(Yes)		1.24 (.21)	0.89	1.73	1.61 (.53)	.85	3.07	1.47 (.38)	0.89	2.44	1.31 (.32)	0.81	2.12	1.63 (.52)	.87	3.05
Academic Attainment		.95*** (.01)	0.92	0.97	.91** (.03)	.86	.97	.91*** (.02)	0.87	0.95	.87*** (.02)	0.84	0.92	.87*** (.03)	.82	.93

Table 3.5. MLM results in primary school sample

Parameter Estimates (Outcome: Key Stage 2)	**Model 1** ***(baseline model)*** Estimate (SE)	**Model 2** ***(Model 1+ aggregate symptom score)*** Estimate (SE)	**Model 3** ***(Model 1+ trajectories)*** Estimate (SE)
Fixed Effects			
Intercept	21.55*** (1.36)	22.21*** (1.36)	22.01*** (1.35)
Prior Attainment: Key Stage 1	.89*** (.02)	.88*** (.02)	.88*** (.02)
Gender (Female)	-.32*** (.09)	-.47** (.10)	-.47** (.10)
FSM (Yes)	-.60*** (.13)	-.55*** (.13)	-.56*** (.13)
SEN (Yes)	-1.11*** (.18)	-1.06*** (.18)	-1.08*** (.18)
Ethnicity (Asian)	.69*** (.20)	.68*** (.20)	.68*** (.20)
Ethnicity (Black)	.13 (.27)	.18 (.27)	.17(.27)
Ethnicity (Mixed)	.01 (.25)	.00 (.25)	.02 (.25)
Ethnicity (Other)	1.20** (.39)	1.16** (.39)	1.12**(.39)
Relative age	-.80*** (.16)	-.81 *** (.16)	-.80*** (.16)
Aggregate externalising symptoms		-.13*** (.02)	-
T2 Low-moderate			-.45*** (.11)
T3 Low-high			-1.29*** (.26)
T4 High-low			-.41*(.19)
T5 High-moderate			-.69*** (.19)
T6 High-high			-.44 (.28)
Variance Components			
Residual variance	2.54 (.03)	2.53 (.03)	2.52 (.03)
School-level variance	1.62 (.12)	1.63 (.13)	1.63 (.13)

Results: 11-14 years

Descriptives

The means, standard deviations and correlations of all the study variables are presented in Table 3.6.

Stage 1: Identifying heterogeneous developmental trajectories

Table 3.7 presents fit indices and criteria for model selection for the single trajectory and the 2-7 class solutions obtained. Based on the entropy, log-likelihood tests and the proportion of cases in the trajectories the 6-trajectory model was selected for further exploration. The 6-trajectory model is presented in Figure 3.3 and Table 3.8 presents sample descriptives and trajectory intercepts and slope co-efficients for each of the 6-trajectories.

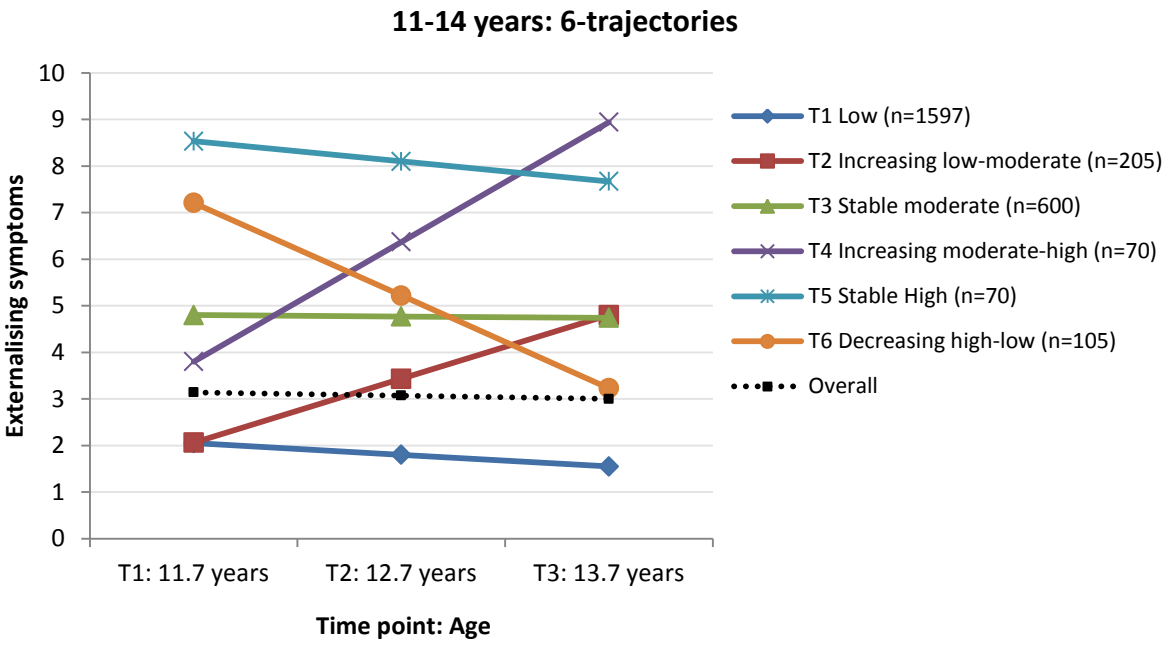


Figure 3.3. Heterogeneous developmental trajectories of externalising symptoms in early adolescents from age 11-14 years (6-trajectory model)

Table 3.8 presents the sample details and the intercept and slope co-efficients for the developmental trajectories in the selected 6-trajectory model and for the overall sample. As can be seen in Table 3.8, a large proportion (60.33%) of adolescents had low externalising

symptom scores in all waves, which is followed by 22.67% with moderate externalising scores in all waves. 7.74% had significantly increasing scores from low to moderate scores and 2.64% had significantly increasing scores into the clinical range (>6). Another 2.64% (N=70) had stable above threshold scores in all waves and 3.97% had significantly decreasing scores from above threshold to low scores.

School level variation in trajectories was estimated to assess the need to account for pupils nesting within schools in further analysis. Amount of school level variation in trajectories was small, ICC=.028 and hence nesting was not accounted for in Stage 2 analysis.

Stage 2: Correlates of trajectory membership

Table 3.9 presents the results of the MLR in the older sample. Gender did not significantly predict having increasing low-moderate symptoms trajectory and being female significantly reduced the probability of belonging to the other trajectory groups when compared to the reference category. Black ethnicity significantly predicted having stable moderate externalising symptoms. Deprivation predicted membership to the stable moderate, increasing moderate-high and decreasing high-low trajectories. SEN did not significantly predict group membership to any of the trajectory groups whereas academic attainment significantly predicted membership to all groups with higher attainment predicting lower probability of belonging to any of the other trajectory groups when compared to the low-symptom reference group.

Stage 3: Predicting academic attainment

Table 3.10 presents the results for the baseline and two follow-up models examining the impact of aggregate symptoms and trajectories of symptoms on subsequent attainment while controlling for prior attainment.

In the older sample, the baseline model indicates that when controlled for prior attainment, individuals who are deprived and have special educational needs have significantly lower attainment scores and females and individuals classified as belonging to other ethnic categories have significantly higher later attainment. Model 2 demonstrates the results when an aggregated symptom score is used to predict change in attainment. The negative significant co-efficient indicates that even when taken as an aggregate higher symptoms significantly predict negative change in attainment, although the effect size is small (β = -.05, ES=.06).

Results of the model including the trajectories, Model 3, indicates that all higher symptom trajectories have a negative impact on later attainment when compared to the low symptom reference group. There are similar significant negative effects of having increasing trajectories, whether to high or moderate symptoms (β = -.14 & -.15; ES =.14 & .18). Stable moderate symptoms also predicted similar levels of negative change (β = -.17, ES =.22) as did the decreasing symptom trajectory (β = -.15, ES =.19). The stable high group predicted the most negative impact on subsequent attainment when compared to the low reference group and the effect size indicates this is a large-moderate effect (β = -.39, ES =.53).

Table 3.6. Means, standard deviations and correlations for variables in the 11-14 year old sample

	Mean (SD)/ Percentage	**Correlations**							
		1.	2.	3.	4.	5.	6.	7.	8.
1.T1 Externalising	3.17(2.41)								
2.T2 Externalising	3.03(2.36)	.58**							
3.T3 Externalising	3.04(2.41)	.50**	.59**						
4.Pre-attainment	4.16(.66)	-.19**	-.15**	-.14**					
5.Post-attainment	5.54(.95)	-.24**	-.20**	-.20**	.81**				
6.Age at T1	11.71(.29)	-.03	-.02	.01	.09**	.07**			
7.Gender	54.4	*-.26***	*-.18***	*-.13***	*.004*	*.05**	*.01*		
8.FSM	17.1	*.08***	*.11***	*.08***	*-.17***	*-.18***	*.01*	*.03*	
9.SEN	7.1	*.10***	*.08***	*.05**	*-.28***	*-.26***	*.00*	*-10***	*.06**

**<.05, **<.001; Italicised are non-parametric correlations (Spearman's rho)*

Table 3.7. Fit indices and model selection criteria for the 2-7 class solutions in the secondary school sample

	% in classes	**Latent class probabilities**	**Entropy**	**A-BIC**	**BLRT**	**LMR-LRT**
2-Class	21.08- 78.92	.82&.93	.68	33897.48	210.88***	202.32***
3-Class	9.2 – 79.11	.75-.92	.74	33716.00	195.59***	187.65***
4-Class	2.91-66.87	.72-.90	.73	33667.91	62.20***	59.68
5-Class	1.85-61.84	.73-.90	.75	33596.93	85.09***	81.64*
6-Class	2.65-60.30	.65-.9	.75	33567.72	43.32***	41.56*
7-Class	2.19-43.41	.68-.86	.71	33523.004	58.83***	56.44

<.10, **<.01, *<.001*

Table 3.8. Descriptives and intercept and slope co-efficients for the identified 6-trajectories in the secondary school sample

Trajectory group	**N (%)**	**Gender % (Female)**	**FSM % (Yes)**	**Age M (SD)**	**SEN % (Yes)**	**Attainment M (SD)**	**Intercept**	**Slope (p)**
T1 *Low*	1597 (60.33)	61.43	14.65	11.71 (.29)	5.26	4.24 (.63)	2.05	-.25***
T2 *Increasing low-moderate*	205 (7.74)	62.95	17.07	11.69 (.29)	6.83	4.22 (.58)	2.06	1.37***
T3 *Stable moderate*	600 (22.67)	41.17	19.83	11.71 (.30)	10.50	4.00 (.70)	4.80	-.03 (.86)
T4 *Increasing moderate-high*	70 (2.64)	42.86	30.00	11.72 (.28)	2.86	4.09 (.72)	3.80	2.57***
T5 *Stable High*	70 (2.64)	22.86	24.29	11.72 (.29)	14.29	3.81 (.61)	8.53	-.43 (.26)
T6 *Decreasing high-low*	105 (3.97)	37.14	25.71	11.70 (.28)	14.29	3.87 (.74)	7.21	-1.99***
Overall sample	2637	54.40	17.11	11.71 (.29)	7.10	4.16 (.66)	3.14	-.07**

Table 3.9. Relative Risk Ratios (RR) for the Multinomial logistic regression of the 6-trajectory model in the secondary school sample

Predictors	**T1 *Low***	**T2 *Inc low-mod***			**T3 *Stable moderate***			**T4 *Inc mod-high***			**T5 *Stable high***			**T6 *Dec high-low***		
		RR (SE)	95% CI		RR (SE)	95% CI		RR (SE)	95% CI		RR (SE)	95% CI		RR(SE)	95% CI	
Gender (Female)		1.03 (.17)	.75	1.42	.43*** (.04)	.35	.43	.42*** (.11)	.25	.69	.17*** (.05)	.09	.32	.38*** (.08)	.24	.58
Ethnicity(Asian)		1.16 (.23)	.79	1.70	.88 (.12)	.67	1.16	.52 (.22)	.23	1.18	.77 (.30)	.35	1.67	.89 (.25)	.51	1.56
Ethnicity(Black)		1.45 (.50)	.74	2.84	1.90** (.42)	1.23	2.93	1.04 (.64)	.31	3.49	1.40 (.88)	.41	4.79	.73 (.44)	.22	2.41
Ethnicity(Mixed)		1.39 (.59)	.61	3.17	1.34 (.39)	.76	2.38	1.82 (1.01)	.62	5.40	1.23 (.94)	.28	5.45	.35 (.36)	.05	2.65
Ethnicity(Other)	**Reference group**	.39 (.40)	.05	2.95	.58 (.30)	.21	1.59	.00 (.00)	.00	-	.00 (.00)	.00	-	.00 (.00)	.00	-
Age		.77 (.20)	.46	1.29	1.15 (.20)	.35	.53	1.09 (.48)	.46	2.59	1.70 (.76)	.70	4.10	1.01 (.36)	.50	2.04
FSM (Yes)		1.16 (.25)	.77	1.76	1.32* (.18)	1.01	1.72	2.18** (.65)	1.22	3.19	1.59 (.51)	.85	2.97	1.98** (.49)	1.22	3.20
SEN (Yes)		1.53 (.49)	.81	2.87	1.19 (.24)	.80	1.75	.15^ (.16)	.02	1.16	1.32 (.54)	.59	2.94	1.44 (.49)	.74	2.80
Academic Attainment		1.06 (.14)	.81	1.38	.60*** (.05)	.51	.70	.62* (.12)	.42	.92	.46*** (.08)	.32	.65	.51*** (.08)	.38	.70

Table 3.10. Multi-level models predicting academic attainment (KS3) post the three waves

Parameter Estimates (Outcome: Key Stage 3)	**Model 1** ***(baseline model)*** Estimate(SE)	**Model 2** ***(Model 1+ aggregate symptom score)*** Estimate (SE)	**Model 3** ***(Model 1+trajectories)*** Estimate (SE)
Fixed Effects			
Intercept	1.15** (.41)	1.26** (.40)	1.25** (.40)
Prior Attainment: Key Stage 2	1.12*** (.02)	1.10*** (.02)	1.10*** (.02)
Gender (Female)	.06** (.02)	.02 (.02)	.02 (.02)
FSM (Yes)	-.13*** (.03)	-.11*** (.03)	-.11*** (.03)
SEN (Yes)	-.17*** (.04)	-.17*** (.04)	-.16*** (.04)
Ethnicity (Asian)	.04 (.04)	.03 (.04)	.03 (.04)
Ethnicity (Black)	.06 (.06)	.06 (.05)	.07 (.05)
Ethnicity (Mixed)	.06 (.06)	.06 (.06)	.06 (.06)
Ethnicity (Other)	.21* (.10)	.19^ (.10)	.17^ (.10)
Relative Age	-.02 (.03)	-.02 (.03)	-.02 (.03)
Aggregate Externalising symptoms		-.05*** (.01)	
T2 *Increasing low-moderate*			-.14*** (.04)
T3 *Stable moderate*			-.17*** (.03)
T4 *Increasing moderate-high*			-.15*(.06)
T5 *Stable high*			-.39*** (.06)
T6 *Decreasing high-low*			-.15** (.05)
Variance Components			
Residual variance	.50 (.01)	.49 (.01)	.49 (.01)
School-level	.25 (.04)	.26 (.04)	.26 (.04)

Discussion

The current study first identified heterogeneous person-centred trajectories of externalising symptom development from ages 8 to11 years and 11 to 14 years. External socio-demographic correlates of different types of developmental trajectories were examined. These trajectory groupings were then used as predictors of change in academic attainment over similar time frame as represented by the symptom development trajectories.

Heterogeneous trajectories of symptoms

In both age groups the largest proportion of individuals had low symptoms over all three measurement points, 49% of the younger sample and 60% of the older sample. This finding is in line with findings from existing studies of externalising symptom trajectories over larger periods of childhood and adolescence, that have found that 40-65% of young people never develop more than low symptoms (egs. Barker & Maughan, 2009; Odgers et al., 2007). Based on evidence from longitudinal studies and Moffitt's taxonomy of externalising symptom development, externalising symptoms are more prevalent and are demonstrated by larger numbers of early adolescence which corresponds to the adolescent limited type of externalising behaviours. In the present study, in line with these theoretical models and previous research findings, nearly a third of the sample had increasing externalising symptom trajectories from age 8 to 11 years. Conversely, the proportion of individuals who develop symptoms in the older sample was comparatively lower (<10%), however almost a quarter of this age group had moderate or high stable symptoms, which corresponds to symptoms present for adolescent limited and persistent problems (Barker & Maughan, 2009). In terms of decreasing symptom trajectories, in the younger sample 13.6% (6.6% high-low, 7% high-moderate) demonstrated trajectories with significant decrease in symptoms, corresponding to the childhood-limited problem group (e.g., 14.7%, Barker & Maughan, 2009). In the older sample the proportion of individuals with decreasing symptoms was low (<4%), and is also explained within the taxonomy of externalising disorders as early adolescence corresponds

with the period of 'maturity gap' where higher numbers are expected to demonstrate symptoms, rather than have decreasing symptoms as that would be expected to occur in middle-late adolescence (Moffitt, 1993). Hence, the results of this stage of analysis identified shorter term developmental trajectories of symptoms that could be placed theoretically within the existing, widely-used taxonomy of development of externalising symptoms through childhood and adolescence.

Correlates of externalising symptom trajectories

Gender was a significant predictor of almost all higher symptom trajectories when compared to the low symptoms trajectory in both age groups. Being male significantly increased the probability of having any higher symptom trajectory in the younger sample and in the older sample males were more likely to have higher symptoms including stable symptoms, with the exception of the increasing low-moderate symptoms which was not significantly predicted by gender. This is consistent with most studies that have found that being male increases the probability of having higher symptom trajectories (Moffitt & Caspi, 2001).

Deprivation significantly predicted membership to the increasing low-moderate trajectory, the decreasing high-moderate symptom and the stable high symptom trajectories in the younger sample, with deprived children being ~2.5 times more likely to have stable high symptoms than non-deprived children. In the older age group, deprivation predicted higher probability of having stable moderate symptoms, decreasing symptoms and increasing moderate-high symptoms. It is possible that the higher probability of belonging to the decreasing symptom trajectory in this sample, which only consists of ~4% of the sample might be due to finding in the younger sample that more deprived children develop symptoms and have high symptoms up to age 11 years. In the older sample, from the 2.64% of individuals with an increasing moderate to high symptom trajectory 30% were eligible for FSM which represents more than twice the relative risk of deprivation predicting increasing

symptoms in this age group when compared to individuals from less deprived backgrounds. The finding that deprivation predicts greater risk of higher symptom trajectories is consistent with existing literature (e.g., Moffitt & Caspi, 2001; Odgers et al., 2007) especially the links with the persistent symptom trajectory, which has been consistently found in the literature.

Ethnicity groupings did not significantly predict the different heterogeneous symptom trajectories in most cases with a few exceptions. Black ethnicity predicted increased probability of having an increasing low to moderate symptoms in the younger group and stable moderate symptoms in the older group. This finding is supported by previous studies of ethnicity and anti-social behaviours in the UK (Smith, 2005), which have found that individuals from Black ethnic groups tend to demonstrate significantly higher externalising behaviours. Existing studies have also found that, in the UK Asians tend to have lower rates of externalising behaviours (Smith, 2005), which was not identified in the present study. Reasons and mechanisms for the differences by ethnicity are still unclear (Nikapota & Rutter, 2009), and are not entirely explained by social disadvantage nor racial discrimination as other minority groups experience both and findings are seen even when controlling for these factors. It has been suggested that more detailed comparisons and examinations of ethnicity and causes, course and correlates of symptom development are necessary to help determine where valid differences lie (Rutter, 2007).

SEN and relative age did not significantly predict membership to any of the externalising trajectory groups in both cohorts. The findings for relative age are difficult to place in existing literature as very little is known about the impact of relative age on externalising symptoms. However, in regards to SEN studies have previously found that individuals with learning disability have higher rates of externalising disorder (Prior et al., 1999), the mechanisms of when the increase in symptomology takes place are not clarified by the results of the current study.

Predicting change in attainment

Analysis assessing the impact of externalising scores on change in attainment scores was carried out using both- a more classical approach of aggregated symptom scores and the derived trajectories from the previous stages. In both age groups aggregate externalising symptom scores significantly predicted negative change in academic attainment, however, the co-efficient was not large nor very informative regarding differential developmental impact as it only suggests impact of amount of externalising difficulties on change in attainment. Analysis using the trajectory memberships as a predictor of change in attainment demonstrates that more differentiated effects of different developmental trajectories could be obtained as compared to the low-symptom group. In the younger sample, all higher symptom trajectories predicted significant negative change in attainment compared to the low-symptom reference group. However, the magnitude and effect sizes indicate that the increasing low-high symptom trajectory had a moderately sized negative impact on attainment, whereas the impact of the low-moderate increasing, high-low and moderate decreasing and stable high symptoms are smaller effect sizes. In the older sample from key stage 2 to key stage 3 assessments, similar to the younger group, all higher symptom trajectories significantly predicted more negative change in attainment scores when compared to the low symptom trajectory group. In this age group increasing symptoms predicted negative impact on attainment, although the effect size was small. Stable moderate symptoms and decreasing high-low symptoms also predicted small effect sized negative impact on change in attainment. However, stable high symptoms over the three waves predicted negative impact on subsequent attainment, controlling for prior attainment. The effect sizes for these effects were moderate-large.

The results support the adjustment erosion hypothesis (Moilanen et al., 2010) and indicate that any higher symptom trajectory results in more negative change in attainment when compared to individuals with no or low symptoms over childhood and early

adolescence. This is consistent with expectations from systemic theories that predict negative changes in one domain to impact negatively on other domains of functioning (Masten et al., 2005).

These results are incongruent with one of the main tenets of Moffitt's (1993) descriptions of symptom development in adolescence. She theorises that externalising symptoms during the maturity gap years in early adolescence are an adaptive development that helps socialisation and have limited negative consequences, and hence posits that symptoms in this maturity gap period are not psychopathology (Moffitt, 1993). In contrast with this prediction, in these data increasing symptoms in this age group are associated with worse academic outcomes. However, we cannot study in these data which of those individuals have truly adolescent limited symptoms as this cohort is limited until age 14 years. Notwithstanding the arguments for whether the symptoms be considered psychopathology or not, the current results demonstrate a significant negative impact of developing externalising symptoms in early adolescence on subsequent academic attainment in nationally mandated tests.

Strengths and limitations

Aside from the methodological strengths already outlined, the use of community based sample in two age cohorts is a particular strength of this study. Studying developmental trajectories in clinical populations alone would not only have made it impossible to include increasing trajectories of individuals who start with low levels of problems, but would also have limited our ability to gain knowledge of population-based estimates and risk factors associated with disorder. Moreover, clinical samples are often subject to referral bias, which results in low generalisability of results (Maughan & Rutter, 2009). The identification and prediction of sub-clinical levels of problems is useful as even though these individuals might not reach specialist services or require resources at this stage information about their risk of developing symptoms can be estimated, especially as sub-threshold levels of problems in

childhood and adolescence increase risk of developing disorder later in life (Fergusson, Horwood, Ridder, & Beautrais, 2005). The inclusion of two age groups (8–11 years and 11–14 years) makes possible comparisons of different risk factors and longitudinal associations with academic attainment at different developmental periods.

In terms of the participants in the study, though the final sample is large and nationally representative, the selective attrition must be noted as a limitation. Participants lost to follow ups were more likely to be deprived, have SEN and lower attainment. The results of many studies, including the present study indicate that these individuals might have had higher risk of having higher levels of problems (Green et al., 2005) and hence the proportions of children with high stable symptoms or increasing symptoms might be underestimated.

Implications and future directions

These results contribute to understanding the nuances in the longitudinal developmental ramifications of externalising symptom development on academic outcomes and adds to existing knowledge of these relationships by permitting a breakdown of the differential effects of different developmental symptom trajectories on development of educational learning.

In terms of size of effects, the current study clearly indicates that increasing symptoms during late childhood and stable symptoms during adolescence have a significant, moderate effect impact on academic attainment. In terms of correlates, the study clarifies that relative age is not a risk factor in the development of externalising symptoms. There is a lack of studies which focus on school based risk factors for symptom development and most of the studies to-date that have examined longitudinal trajectories of symptoms have focussed on family, neighbourhood and individual characteristics (e.g., Barker & Maughan, 2009). Considering the importance of the school environment to children's development (Eccles &

Roeser, 2011), the study of classroom and school level risk factors in the development of externalising symptoms might lead to some interesting results.

The study of person-centred trajectories of externalising symptoms is widely established and the current study mainly makes a contribution to existing literature in this area in terms of the impact of shorter term developmental patterns of externalising symptoms on change in attainment in these two cohorts. However, in contrast, the study of internalising symptom development is more recent and does not have well-established theory or consistently identified groupings of trajectories through childhood and adolescence. Moreover, the longitudinal links between internalising symptoms and academic performance have been difficult to establish, with decades of inconsistent results being reported in the literature at all ages and in different samples (Masten et al., 2005). Using a similar methodology as in this study, the next study examines the development of internalising symptoms, its correlates and its impact on change in academic attainment.

Study 4. Developmental trajectories of internalising symptoms: patterns, correlates and links with academic attainment

Background

Internalising symptoms refer to symptoms of disordered mood or emotion (Kovacs & Devlin, 1998), and the term is widely used in child psychopathology research as an umbrella term including the various symptoms of depression and anxiety(Goldberg & Goodyer, 2005). Disorders with internalising symptoms are one of the largest forms of mental health problems faced by individuals of all age groups: children, adolescents and adults (Lewinsohn et al., 1993; Patel, Flisher, Hetrick, & McGorry, 2007). In the developed world they are the second highest cause of DALY's (Murray et al., 2012) and the WHO predicts depression will be the number one cause of disability by 2030 (WHO, 2008). Internalising symptoms, aside from being one of the main mental health problems faced in childhood and adolescence (Green et al., 2005), are a strong predictor of developing a diagnosis in adulthood (Roza, Hofstra, van der Ende, & Verhulst, 2003). Studies have investigated longitudinal change in symptoms to help better understand development of psychopathology (Costello et al., 2011) and have helped identify that internalising symptoms generally increase as children become adolescents (Costello et al., 2011) and decrease as adolescents move into adulthood (Galambos, Barker, & Krahn, 2006).

The problems associated with emotional disorders can have a multitude of effects on individuals, their development and relationships with friends, family and the wider society. These effects could manifest through a variety of channels. For instance, there is evidence that depressive people generate negative responses from people they interact with, including negative emotional and verbal/non-verbal responses which are indicators of rejection in the form of lack of desire for further contact (Coyne, 1976; Gotlib & Beatty, 1985; Gotlib & Meltzer, 1987). Depressive symptoms in childhood have been associated with insecure attachments in childhood (Cummings & Cicchetti, 1993), dysfunctional family relationships (Kaslow, Deering, & Racusin, 1994), maternal/paternal psychopathology (Ramchandani & Psychogiou, 2009) and difficulties in peer relationships (Panak & Garber, 1992).

Considering the prevalence and impact of internalising symptoms, examining the course of symptoms over time can offer invaluable clues to the aetiology, risk and protective factors and development of symptoms and disorder. With the increasing availability of longitudinal data, symptom development has been the focus of much research.

Longitudinal development of internalising symptoms

In the largest epidemiological study of prevalence of mental health problems in young people in Britain, BCAMHS (Green et al., 2005) rates of anxiety and depressive symptoms were significantly higher in adolescence, with rates around 3.5% in 5-7 year olds and 7-8% in 13-15 year olds (Ford et al., 2003). A review of epidemiological studies by (Costello et al., 2011) looking at changes from childhood to adolescence demonstrated that most anxiety and depressive diagnoses increase when children become adolescents with the exception of separation anxiety disorders and in some studies specific phobias (ibid). Studies also suggest that depressive symptoms decrease overall as adolescents move into the phase referred to as emerging adulthood (18-25 years) (Galambos et al., 2006). Hence emotional symptoms peak in the first three to four years of early/middle adolescence (Poulin, Hand, Boudreau, & Santor, 2005). Studies also indicate that adolescent problems are a pre-cursor for problems during adulthood (Roza et al., 2003), with sub-clinical levels of problems also predicting greater risk of diagnosis in adulthood (Fergusson et al., 2005).

Conventional approaches have used a 'single-trajectory' approach to modelling change which rests on the assumption that as individuals come from a single population a single growth trajectory adequately captures change (Jung & Wickrama, 2008). Although in overall growth trajectory approaches individual slopes are allowed to vary from the overall trajectory, this variation is difficult to capture in assessing the correlates of the slope and the impact of the slopes. Hence, the homogeneity assumption of these techniques encompasses the assumption that covariates that affect growth influence each individual the same way. Contrary to this assumption, findings from several research papers looking at growth models

over time in areas such as conduct problems (Roisman et al., 2010), alcohol use (Greenbaum et al., 2005; Jackson & Sher, 2005), antisocial personality disorder (Bucholz, Hesselbrock, Heath, Kramer, & Schuckit, 2000) etc. indicate that there is a heterogeneity of growth trajectories and that using a single growth trajectory estimate is an over-simplification. Although the importance of studying individual trajectories has been acknowledged (e.g., Sroufe & Rutter, 1984), in child and adolescent emotional psychopathology research the focus has mostly been on the risk factors and predictors of problems (variable-centred approach) rather than the possible heterogeneity of individual trajectories over time (person-centred approach). In the absence of a single population trend for change in emotional symptoms across childhood and adolescence existing single-trajectory estimates make it difficult to understand how subgroups of individuals change as it can obscure distinct differences in how subgroups have different patterns of change (Sterba et al., 2007; von Eye & Bergman, 2003). The possible pitfalls of assuming a single homogenous trajectory of emotional symptom is illustrated by mixed findings of the single trajectory approaches from around the same age range (childhood-pre-adolescence) with some studies finding increased symptoms over these years (Gazelle & Ladd, 2003), some stable (Keiley et al., 2003) and some decreasing (Green et al., 2005, Wolpert et al., 2011). In reality the likelihood is that the average trajectory includes some decreasing, increasing and stable trajectories and the sample mean reflects the largest grouping. Studies in other domains of development, especially those focusing on behavioural problems, have been using a multi-trajectory approach to understand problem behaviours and this accounting for heterogeneous growth trajectories has improved the understanding of the different patterns and risk factors of behavioural problems in young people (e.g., Roismann et al., 2010).

Studying the heterogeneous trajectories of internalising symptoms has gained popularity in the last decade as it allows researchers to partition the effects of variables and time on different individuals who experience varying patterns of symptom development.

These person-centred approaches marked a substantial leap forward in our understanding of aetiology and development of psychopathology as they dropped the prior help assumption that the relationships across predictors, time and individuals was constant (von Eye & Bergman, 2003).

Studies have identified the predictors associated with trajectories of both the broader domain of internalising symptoms (Côté et al., 2009; Sterba et al., 2007); and those of the main sub-types: depression (Dekker et al., 2007) and anxiety (Broeren, Muris, Diamantopoulou, & Baker, 2013; Crocetti, Klimstra, Keijsers, Hale, & Meeus, 2009; Legerstee et al., 2013; Morin et al., 2011). Some studies have focussed on the development of symptoms in early childhood (1-5 years, Cote et al., 2009; 1-3 years, Carter et al., 2010), which explore symptom development and find different trajectories of symptoms starting from low to low, moderate or higher symptoms over this young age. Dekker et al. (2007) explored trajectories from age 4-18 years and explored six trajectory models for both males and females. Sterba et al. (2007) investigated internalising trajectories from 2-11 years and identified three distinct trajectories in each gender and examined their correlates. Hence, as can be seen, unlike with externalising symptoms (e.g., Roisman et al., 2010, detailed in introduction to Study 4), there isn't a clear taxonomy of internalising symptom trajectories with supporting theoretical and empirical evidence. Hence, research investigating internalising symptom trajectories through childhood and adolescence is driven more by data and exploratory analyses in existing data-sets. Studies of short-term heterogeneous trajectories of emotional symptoms in school age children are rarer with only one study identified, Broeren et al. (2013) which explored trajectories over three waves in 224 children ranging from 4-9 years at the initial wave.

Although these studies have contributed greatly to the study of the development of internalising symptomology in childhood and adolescence most of these studies have in common the use of 1) a wide age range (e.g., from 4-18years, (Dekker et al., 2007)), 2) the

use of proxy reported symptoms and 3) the lack of focus on predictive impact on other domains of childhood functioning. The only educational outcome that has been explored by a few studies is level of education attained by adulthood (Dekker et al., 2007; McLeod & Fettes, 2007) and although this is interesting and important, there is no research that has focussed on the shorter term impact of increases and decreases in symptomology on attainment.

Correlates of internalising symptoms

An important aspect in studying epidemiological patterns and mechanisms has been identifying the various factors, including socio-demographic, that influence the risk of developing disorder (Verhulst & Koot, 1992). Examining the predictors of development of symptoms over time offers invaluable clues about different individuals' differential risk of developing, maintaining and recovering from symptoms.

Studies in the past have looked at predictors of the whole-population trajectory. This approach masks individual variations in change and assumes a common relationship between symptoms, time and predictors for all individuals in a population; this gives results that can be atypical of any individuals in a population, and hence, are not very useful (von Eye, 2010). Studies that have identified heterogeneous independent symptom trajectories have mainly focused on gender, alongside parental variables such as maternal depression, as predictors of different developmental trajectories (Côté et al., 2009; Sterba et al., 2007).

The current study will include gender along with other socio-demographic and educational predictors such as ethnicity, deprivation, relative age, special educational needs and educational attainment. These variables are of interest because they 1) are associated with internalising symptoms in cross-sectional studies (Goodman et al., 2003; Green et al., 2005), 2) predict different trajectory memberships in the other key child mental health domain, conduct problems (Roisman et al., 2010) and, 3) predict other key childhood

outcomes such as academic attainment (Cole, Martin, Powers, & Truglio, 1996) and risky health behaviours (Brooks, Harris, Thrall, & Woods, 2002).

Gender. Gender is a well-established predictor of emotional problems in adolescents and adults (Hankin et al., 1998; Nolen-Hoeksema, 1987; Nolen-Hoeksema & Girgus, 1994), for e.g., in adulthood females are twice as likely as males to have depression (Nolen-Hoeksema, 1987). Even in pre-adolescents, some studies have found that females have higher amounts of emotional symptoms than males (Green et al., 2005) and there are studies showing no differences preadolescence (e.g., (Fleming, Offord, & Boyle, 1989) and higher rates in pre-adolescent boys (Anderson, Williams, McGee, & Silva, 1987; Angold, Costello, & Worthman, 1998). However the majority of research in children indicates that gender differences in emotional disorder appear or widen during or immediately post-puberty (e.g., Angold et al., 1998). (Wade, Cairney, & Pevalin, 2002) explored three different national long datasets to determine when the gender gap arises and concluded that the gender gap in depression manifests itself in adolescents by the age of 14, when onwards females exhibit higher levels of depressive symptomology into adulthood. Angold et al. (1998) in the GSMS also identify between 10-15 years as when the gender gap emerges, however they suggest that physical maturity is a better predictor than chronological age. Some studies have also indicated that it is only females who have increased risk of developing emotional diagnosis in adolescence (Costello, Mustillo, et al., 2003).

Studies of heterogeneous trajectories have examined gender differences in development of symptoms and discovered that the same number of heterogeneous developmental trajectories best summarise symptom development over time in the same sample (Dekker et al., 2007; Sterba et al., 2007), although they have slightly different developmental trends. Dekkar et al. (2007) examined some young adult outcomes predicted by these trajectories and conclude that although the trajectories differ by gender, the young adult outcomes are not very distinguishable for the different high trajectory symptoms.

SES. Deprivation or low SES have well-established links to mental health problems in childhood (Green et al., 2005). Some studies (Gilman, Kawachi, Fitzmaurice, & Buka, 2002) have shown that low SES children are twice as likely as having experienced depression when compared with high SES children. As lower SES is often an indication of ongoing adverse social and environmental conditions, it can be considered a continuous or chronic stressor (Baum, Garofalo, & Yali, 1999). Sadowski, Ugarte, Kolvin, Kaplan, & Barnes, (1999) found that deprivation in childhood also predicted depression in adulthood. In a longitudinal perspective, (Rushton et al., 2002), in a US community sample found that SES (and ethnicity) did not predict persistent depressive symptoms over one year in 12-17 year olds.

Ethnicity. Unlike with externalising symptoms, links between socio-ethnic groupings and development of emotional symptoms have not been clearly established (Twenge & Nolen-Hoeksema, 2002), one reason being that they have not been studied as much (Nikapota & Rutter, 2009). Cross-sectional explorations suggest that some ethnic minority groups might have higher levels of emotional disorders, (Roberts & Cawthorpe, 1995) for example found that Asians in Britain were more likely to have emotional disorders and receive specialist treatment for these problems. Possible explanations include differences in genetic vulnerability, cultural differences and socio-economic position (Nazroo, 1998). Importantly, across ethnic groups studies have shown the broad risk and protective factors against disorder remain the same (neglect increases risk; continuous good family relationships are protective (Nikapota & Rutter, 2009). However, studies indicate that there are socio-ethnic variations in which some mechanisms operate, where certain factors might be a moderator in one and not in another ethnic group. For instance, parenting styles mediate the risk effects of maternal depression on child mental health in White and Hispanic samples but do not in Black samples in the US (Pachter, Auinger, Palmer, & Weitzman, 2006). Hence, other hypotheses for ethnic differences include cultural and familial institutions providing a protective or risky environment which can decrease or increase risk of certain disorders.

Relative age. As detailed in the background to the previous study, no studies were identified that examined the impact of relative age specifically on externalising symptoms or anti-social behaviours in children. One study in Britain, (Goodman et al., 2003) investigated overall total difficulties and found that younger children in cohorts have significant overall difficulties; however, the effect on sub-domains of child psychopathology and their development in this age-group has not been investigated.

SEN. Children with SEN are at greater risk of exclusions, absenteeism and poorer educational outcomes in schools in England (DFE, 2010a). Studies also indicate that they have worse psycho-social outcomes (Humphrey, Lendrum, Barlow, Wigelsworth, & Squires, 2013) and are more likely to experience bullying (Van Cleave & Davis, 2006). Specifically, studies have shown that children with intellectual disability experience higher internalising symptoms (Emerson, 2003), although their associations with symptom development trajectories has not been explored.

Internalising symptoms and academic attainment

Evidence for the longitudinal associations between internalising symptoms and educational attainment is sparse and results of existing studies are largely inconclusive (Masten et al., 2005). However, cross-sectional associations between internalising symptoms and educational attainment are more established (Fauber, Forehand, Long, Burke, & Faust, 1987; Fröjd et al., 2008) with higher levels of symptoms being cross-sectionally associated with greater academic problems. The difficulty in establishing longitudinal associations in some part might be attributed to the complex nature of internalising symptoms and symptom development (Fergusson, Horwood, & Boden, 2006; Masten et al., 2005), and in other parts to possible differences between depression and anxiety symptoms in their associations with attainment confounding attempts to examine longitudinal associations. However, results are inconclusive with both anxiety and depressive symptoms, although a recent review suggests that the evidence for the negative impact of depressive symptoms is supported by a few

studies (Riglin, Petrides, Frederickson, & Rice, 2014). In terms of anxiety symptoms, there are studies suggesting a non-linear relationship with moderate sub-clinical levels of symptoms being linked with better academic performance when compared to high anxiety (DiLalla, Marcus, & Wright-Phillips, 2004; Sharma, 1970), whereas other research indicates that symptoms relate to poor performance (Gumora & Arsenio, 2002; Weeks, Coplan, & Kingsbury, 2009). Hence, the lack of consistency in results is observed across different types of internalising disorders and is unlikely to be explained by the depression-anxiety differences alone, especially considering the two sub-domains of internalising symptoms are intrinsically linked and highly co-morbid during childhood and early adolescence (Moffitt et al., 2007).

Methodologies used to explore these longitudinal associations have included: an early symptoms predicting later outcomes approach (Duncan et al., 2007; Fergusson & Woodward, 2002), aggregate symptoms predicting change in educational outcomes (Cole et al., 1996), and more complex approaches such as cascade modelling (Masten et al., 2005; Moilanen et al., 2010; Vaillancourt et al., 2013); with some of the larger studies suggesting no link between symptoms and later educational attainment (Cole et al., 1996; Duncan et al., 2007). Cole et al. (1996) using an aggregate symptoms predicting subsequent attainment approach found no significant effects of internalising symptom levels in a large sample of American adolescents. Similarly, Moilanen et al. (2010) conducted cascade analysis and found no links between earlier internalising symptoms with later academic competence. (Romano, Babchishin, Pagani, and Kohen 2010; Duncan et al., 2007) also examined these relationships in large datasets across different countries (Canada, US, UK) and did not find significant links between emotional symptoms to later educational attainment. Conversely, links have been found between adolescent levels of depression and education attained at age 21 (Bardone, Moffitt, Caspi, Dickson, & Silva, 1996) and negative impact of internalising symptoms in early adolescence on subsequent attainment within the same academic year (Riglin et al., 2013).

The approach to examining this relationship used in existing studies have in common the limitation of grouping together all individuals in the sample which makes the assumption that the relationship between time, symptoms and attainment across all individuals is the same. A trajectory based approach by Dekkar et al. (2007) who examined the relationship between trajectory membership from five waves of assessment from age 4-18 years by reporting percentages in each trajectory who attained a low level of education in adulthood and found that males with chronic symptoms throughout childhood and adolescence were more likely to have lower educational level attained by adulthood. No study was identified that has assessed the differential impact of heterogeneous symptom trajectories during childhood and adolescence on immediately subsequent academic attainment outcomes while controlling for prior attainment to allow assessment of the impact of differential symptom development on change in attainment over a similar period as represented by the symptom trajectory.

The current study

The current study aims to contribute to understanding shorter term development of symptoms in late childhood and early adolescence (8-14 years) by identifying heterogeneous trajectories of internalising symptoms in two community based longitudinal cohorts, aged 8-11 years and 11-14 years. Once identified, the socio-demographic (gender, ethnicity, deprivation and relative age) and educational (special educational needs [SEN], academic attainment) predictors of different symptom trajectories will be explored to identify unique associations between risk factors and different symptom pathways. Subsequently, different symptom trajectories will be compared to symptom severity aggregated across time as a predictor of change in educational attainment over the same time frame. This will not only help assess the predictive utility of using a heterogeneous approach, but also potentially provide a unique approach to help disentangle the longitudinal effects of emotional symptomatology on educational outcomes.

The unique contributions of the current study relate to 1) employment of two cohorts in primary and secondary schools (aged 8-11 years & 11-14 years) which allows comparisons of trajectories, their predictors and how they predict change in educational attainment in the two age groups, 2) exploring the socio-demographic correlates of these short term developmental trajectories and, 3) using trajectory membership as a predictor of change in another key outcome, educational attainment.

It is hypothesised that internalising symptom development across three waves in late childhood, early adolescence will be better explained by heterogeneous trajectories, than a single trajectory. Secondly, different symptom trajectories are expected to impact differentially on change in educational attainment over the same period.

Method

Design

The design of this study is similar to the previous study of externalising symptoms. Internalising symptoms were assessed in the first term of schooling (which is during autumn in England) every consecutive year for three years. Educational attainment scores were taken from national standardised tests at the end of a Key Stage (KS) in England which correspond to age 7 (KS1), age 11(KS2) and age 14 (KS3) (see www.education.gov.uk/ for detailed descriptions of KSs). As a result participants had a KS score prior to the start of the study and a KS score post the three waves (see Figure 4.1). The manner in which the two educational attainment measurement points frame the three waves of data pertaining to participants' internalising symptoms provides a scenario where development of symptoms during the three waves can be used to examine the impact of development of symptoms on relative change (gains or losses) in national standardised tests of educational attainment.

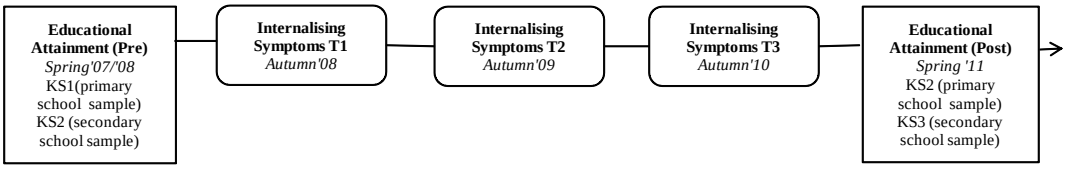


Figure 4.1. Design of the study illustrating when data for key variables, educational attainment and internalising symptoms, were collected

Participants

The participants in this study are the same as in the previous study and at the first wave of data collection the mean age was 8.70 years (SD = 0.30) in the primary and 11.71 years (SD = 0.29) in the secondary school sample. The sample, it's generalisability and the attrition details are detailed in the previous study (pages 119-121).

Specifically in terms of internalising symptoms the final primary sample did not have significantly different scores from the wave 1 sample ($t=1.66$, $p=0.096$) and in the secondary school sample the final sample had significantly lower mean scores when compared to the wave 1 sample ($t=2.34$, $p<.05$).

In summary, the final sample analysed in this study is representative of national pupils except for deprivation, where the study sample has a slightly higher proportion of deprived pupils. However, attrition indicates that pupils lost in follow up waves were significantly more likely to be FSM eligible, have special educational needs and have lower attainment scores.

Procedure

The procedure for data collection in this study was the same as in the previous study (Study 3, see page 119)

Measures

Internalising symptoms

Internalising symptoms were measured using the emotional difficulties scale of the Me and My School questionnaire (M&MS; Deighton, Tymms, et al., 2013). The scale is a 10-item self-report measure comprising of items such as 'I feel lonely' and 'I worry a lot' (see Appendix B for list of all items). Participants respond to each item by selecting one of three options: Never, Sometimes, Always. The answers to the items are summed to create a total emotional difficulties score, higher indicating more difficulties. The scale has an at-risk cut-off score of 10 (ibid.).

Academic Attainment

The measures of academic attainment pre and post the three waves of externalising measures is the same as used in the previous study (see Page 122 for details). In summary, for the primary school sample, the KS1 score (M = 15.12; SD = 3.47) and the KS2 score (M = 27.76; SD = 4.21) were used as measures of attainment pre and post the three waves. In the secondary school sample, KS2 levels (M = 4.16, SD = 0.66) and KS3 levels (M = 5.54, SD = 0.96) were used as measures of attainment pre and post the three waves. **.**

Socio-demographic correlates

The socio-demographic correlates in this study are identical to the previous study (gender, ethnicity, SES, SEN and age) and are described in detail on pages 122-123.

Analytic strategy

Analyses were conducted in 3 stages to answer the research questions discussed above. The approach is similar to the previous study and hence described in only brief below (for rationale and more detailed description please see previous study pages 123-129).

Stage 1: Identifying heterogeneous developmental trajectories

Latent Class Growth Analysis (LGCA) was conducted using MPLUS (Version 7, Muthén & Muthén, 2012) to identify a *k*-class model that had good fit criteria and interpretability. Two- class to seven-class models were explored in both the primary and secondary school samples.

Criteria used to select the k-class model for further analyses included assessing 1) proportions identified in classes, 2) neatness of classification (measured by entropy and posterior probabilities, Jung & Wickrama, 2008), 3) information criteria, due to the large sample size sample size adjusted BIC was used (Yang, 2006), 4) likelihood ratio tests comparing a model with k classes with a k-1 class model to determine if the k model is significantly better that the k-1 class model. The Lo-Mendell-Rubin likelihood ratio test (LMR-LRT) and bootstrapped likelihood ratio test (BLRT) were both estimated.

Stage 2: Correlates of heterogeneous symptom trajectories

Multinomial logistic regressions (MLR) were conducted in STATA 12 (StataCorp, 2011) to determine the socio-demographic and school-related predictors of trajectory membership when compared to the reference group (as per convention the group with the largest proportion of participants was selected to be the reference category). Relative risk ratios (RR) that represent the probability for the predictor of interest of having a certain trajectory when compared to the reference group were estimated to allow for easy interpretation (McNutt et al., 2003). A RR greater than 1 indicates that the risk is increased for the predictor category/unit change in predictor in question and, inversely, RR's less than 1 indicate reduced risk.

Stage 3 Predicting academic attainment

MLMs were conducted to examine the relationship between the different heterogeneous trajectories of symptom development and the corresponding change in academic attainment over the three waves. This was assessed by analysing trajectory groups

predict attainment after the last wave after controlling for attainment prior to the initial wave. As school-level variation in attainment was high (>20%), multi-level models (MLM) were used to account for nesting. Like in Study 3, MLMs were computed using both aggregated symptom scores over three waves and the derived trajectories as predictors to allow comparison of the predictive utility of the trajectory based approach. In this stage of analysis the following MLM's were computed

4. Baseline model: A baseline model with all socio-demographic predictors was run as preliminary to further analysis
5. Aggregate symptoms: Model 1 with aggregated symptoms over 3 years to assess the extent to which aggregated symptoms across all individuals predict change in academic attainment
6. Trajectories: This analysis includes the baseline model along with the trajectories, coded as categorical variables, predicting attainment. The co-efficients indicate the effect of being in each trajectory group when compared to the reference category.

Results

Results of all three stages of analysis are first presented for the primary school children (8-11 years), followed by the secondary school sample (11-14 years).

Results: 8-11 years

Descriptive statistics

The means, standard deviations and correlations of all the study variables are presented in Table 4.1.

Stage 1: Identifying heterogeneous developmental trajectories

Latent class growth analyses were conducted in Mplus (Muthén & Muthén, 2012) to identify a *k*-trajectory model that had good fit criteria and interpretability. Two- to seven-trajectory models were explored and fit criteria and model estimates are presented in Table 4.2.

As can be seen in Table 4.2 there was not a single model that was the best on all outlined criteria. Based on the different model fit criteria, the 4- and 6- class models were both considered as they had good fit and were significantly better than the next model as indicated by the drop in BLRT and A-BIC. However, the LMR-LRT which is a more conservative LRT indicated that the 5-class model was not a significant improvement on the 4-class model. Based on the LMR-LRT and entropy the four class model seemed to be a good fit, however about 60% of the sample were in the low-steady group which provided very little heterogeneity. The 6-class model was also considered as the fit indices were good and the model was significantly better than the 4-class model (2*log-likelihood difference 4 vs 6 class model = 74.1, $p<.001$), while having 39% of the sample in the largest group. Figures 4.1 and 4.2 demonstrate the trajectories obtained in the 4 and 6 class trajectory models respectively.

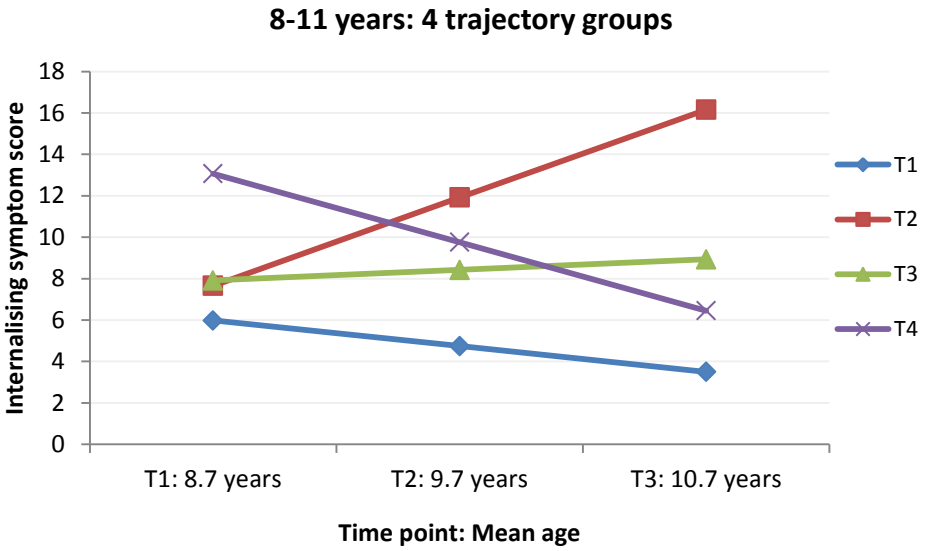


Figure 4.1. 4-trajectory solution in the primary school sample

As can be seen from Figures 4.1 and 4.2, the 6-class model essentially differentiated the large low-low class in 4-class model into three classes; low-low, moderate- moderate and high-high. In research most focus is usually on the groups that decrease from clinical scores or increase to clinical scores over time as it is interesting to see what characterizes individuals with change over time. However, as noted earlier, it is also interesting to study individuals who have steady moderate or steady sub-clinical levels of difficulties as even though these individuals might not reach specialist services or require resources, the predictors and outcomes of these trajectories are interesting as they potentially give us more information on what increases risk of children with sub-clinical levels of symptoms, which may not merit clinical intervention, but are nevertheless of importance as they might adversely affect normal development and functioning in other domains of development (Reinherz, Giaconia, Lefkowitz, Pakiz, & Frost, 1993) and have been shown to predict adult psychopathology (Fergusson et al., 2005). To establish if a more heterogeneous 6-class model gave us more information than the 4-class model, predictors of the classes in the 6-class model were explored with the idea that if the extra classes in the 6-class model had significantly different

predictors/risk factors than the decreasing low trajectory, we would be justified in further exploring the 6-class model for its increased theoretical significance.

Based on the following reasons the six-trajectory model was chosen for further exploration: (1.) heterogeneity; the 6-class model showed greater heterogeneity as the largest group had ~40% of sample and the proportion in the smallest group did not differ between the models (.9%), (2.) theoretical interest; 6-class model identified the moderate stable and high stable trajectories in addition to the 4 already identified, and (3.) an analysis of the predictors (as presented in Stage 2) which indicated that the 2-extra classes identified by the 6-class model had significantly different predictors.

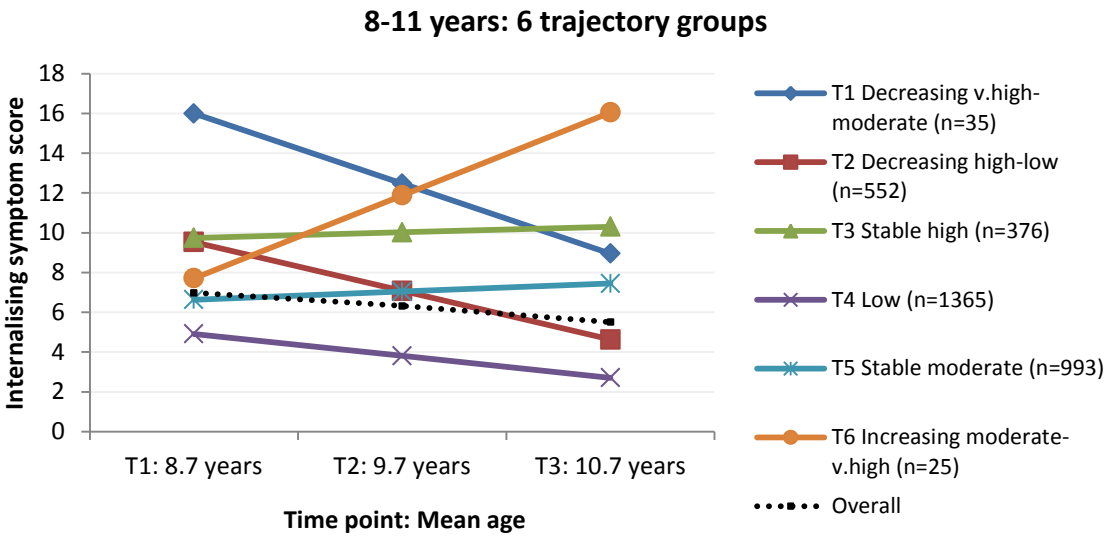


Figure 4.2. Developmental trajectories of internalising symptoms in children aged 8-11 years

Table 4.3 presents the sample details and the intercept and slope co-efficients for the developmental trajectories in the 6-trajectory model and for the overall sample. As can be seen from Table 4.3, proportions of children with the different symptom trajectories over 3 waves varied. As would be expected, the largest proportion of participants had low symptoms over time (T4, 40.8%). Two groups had stable symptoms over time, moderate (29.7 %) and high (11.2 %). The proportion of children with significant increases in scores was low in this sample (less than 1%), which indicates that not many children developed high symptoms over

this time. On the other hand 16.5% had steep decreasing symptoms from high to low symptoms and a small proportion had decreasing high-moderate symptoms (1%).

School level variation in trajectories was estimated to assess the need to account for pupils nested within schools in further analysis. Amount of school level variation in trajectories was small, ICC= .004 hence nesting was not accounted for in the next stage of analysis.

Stage 2: Correlates of heterogeneous symptom trajectories

MLR was conducted with the trajectory with low symptoms at all the time points (T4) as the reference category. Table 4.4 presents the relative risk ratios (RR), along with the standard error (SE) and the 95% confidence intervals (95% CI) of the MLR for the primary school 6-trajectory model. In the younger cohort (see Table 4.4), gender and relative age significantly predicted membership to almost all the higher symptom trajectories, except the increasing trajectory, when compared to the reference group; with females and children younger within their cohort being at significantly higher risk. Being Asian significantly increased risk of being in stable moderate group in the primary sample (RR =1.28, $p<.05$) and Black ethnicity significantly increased the probability of being in the decreasing high-moderate trajectory (RR = 4.51, $p<.01$). FSM increased risk of having stable high levels of symptoms (RR = 1.39, $p<.05$) and SEN increased probability of being in both the decreasing trajectories and the stable high trajectory. Higher academic attainment scores significantly predicted lower risk of being in categories other than the reference group.

Stage 3: Predicting change in academic attainment

To predict change in attainment, MLMs were conducted predicting subsequent attainment while controlling for prior attainment. Table 4.5 shows the results of the MLMs predicting Key stage 2 results in the primary sample, controlling for prior attainment (Key Stage 1). The following models were estimated: the first model (1) controls for prior attainment and socio-demographic correlates, the second model (2) is Model 1 including aggregated internalising symptom score as a predictor and the final model (3) is Model 1 with trajectories added in as categorical predictors; the reference category being the low symptom trajectory group. Effect sizes (ES) were computed by dividing the beta estimate for main effect by the average of the square root of the variance estimate at both time points (Fonagy et al., 2009).

As can be seen in Table 4.5, prior attainment was a strong predictor of attainment and gender, deprivation, ethnicity, relative age and SEN also significantly predicted relative change in attainment over the three years. Results of Model 2 indicate that higher aggregate symptom scores significantly predict negative change in attainment ($\beta = -.06$, $p<.001$, ES=.02). On the other hand the analysis with the trajectories (Model 3) provides a greater breakdown of effects. In the younger sample there are significant negative effects of having a decreasing high-low trajectory ($\beta = -.28$, $p<.05$, ES =.07) and stable moderate symptoms ($\beta = -.35$, $p<.01$, ES =.09). The other three trajectories, decreasing high-moderate, stable high and increasing moderate-high do not predict significantly different attainment scores compared to the reference low symptom group (ES=.05-.06).

Table 4.1. Means, standard deviations and correlations for variables in the 8-11 year old sample

	Mean (SD) *or* Percentage	**Correlations** 1.	2.	3.	4.	5.	6.	7.	8.
T1 Internalising	6.97 (3.47)	1							
T2 Internalising	6.32 (3.56)	.48**	1						
T3 Internalising	5.51 (3.49)	.41**	.56**	1					
Pre-attainment	15.12 (3.47)	-.13**	-.15**	-.13**	1				
Post-attainment	27.76 (4.21)	-.14**	-.16**	-.14**	.72**	1			
Age at T1	8.70 (.30)	-.08**	-.10**	-.09**	.19**	.09**	1		
Gender (Female)	49.3%	*.17***	*.!3***	*.13***	*.09***	*.04***	*-.03*[*]	*1*	
FSM	18.1%	*.03*	*.08***	*.07***	*-.22***	*-.23***	*.01*	*.00*	
SEN	8.6%	*.07***	*.09***	*.09***	*-.34***	*-.31***	*-.06*[**]	*-.14*[**]	*.13*[**]

**<.05, **<.001; Italicised are non-parametric correlations (Spearman's rho)*

Table 4.2. Fit indices and criteria for model selection for 2-7 class LCGA solutions

	BLRT	**LMR-LRT**	**A-BIC**	**Entropy**	**Posterior probabilities**	**Estimated % in classes**
2- Class	156.78**	150.6**	51400.00	.58	.74-.89	20 - 80
3-Class	100.42**	96.46**	51314.39	.62	.68-.85	6.5-68
4- Class	72.57**	69.71*	51256.63	.7	.69-.89	.9 - 59
5-Class	37.22**	35.75	51234.23	.67	.64-.88	.8-54
6-Class	36.99**	35.53	51212.05	.65	.66-.88	.9-39
7-Class	25.79**	24.77	51201.08	.62	.58-.89	.8-39

**<.05, **<.001*

Table 4.3. Sample breakdown, socio-demographic descriptives and intercept and slope co-efficients for the trajectory groups in the primary school sample (8-11 years)

Trajectory group	**N (%)**	**Gender % (Female)**	**FSM % (Yes)**	**Age M (SD)**	**SEN % (Yes)**	**Academic Attainment M (SD)**	**Intercept**	**Slope (p)**
T1 *Decreasing high-moderate*	35 (1)	62.9	25.7	8.60 (.30)	20	13.60 (2.91)	15.99	-3.52***
T2 *Decreasing high-low*	552 (16.5)	57.1	18.3	8.69 (.29)	10.6	14.75 (3.54)	9.53	-2.45***
T3 *Stable high*	376 (11.2)	60.4	24.5	8.67 (.32)	13.6	14.21 (3.66)	9.74	.28(.68)
T4 *Low*	1365 (40.8)	41.2	15.5	8.74 (.29)	6.1	15.64 (3.38)	4.91	-1.1***
T5 *Stable moderate*	993 (29.7)	51.8	19.5	8.68 (.30)	8.7	15.04 (3.35)	6.63	.41 (.26)
T6 *Increasing moderate-high*	25 (.7)	40	24	8.65 (.31)	20	13.94 (3.49)	7.71	4.17***
Overall primary sample	3346	49.3	18.3	8.70 (.30)	8.7	15.12 (3.47)	6.98	- .74***

Table 4.4. Relative Risk Ratios (RR) for the Multinomial logistic regression of the 6-trajectory model in primary school sample

	T1 *Decreasing high-moderate*			**T2 *Decreasing high-low***			**T3 *Stable high***			**T4 *Low***	**T5 *Stable moderate***			**T6 *Increasing low-high***		
	RR (SE)	95% CI		RR (SE)	95% CI		RR (SE)	95% CI			RR (SE)	95% CI		RR(SE)	95% CI	
Gender (Female)	3.02** (1.11)	.47	.20	2.06*** (.22)	1.67	2.54	2.48*** (.31)	1.94	3.17		1.6*** (.14)	1.35	1.9	1.03 (.45)	.44	2.41
Ethnicity-Asian	.79 (.49)	23	.68	.99 (.16)	.72	1.37	1.09 (.2)	.76	1.56		1.28* (.16)	1	1.65	1.78 (.9)	.65	4.85
Ethnicity-Black	4.51** (2.34)	.63	2.50	1.56^ (.38)	.97	2.52	1.68^ (.46)	.98	2.85		1.36 (.29)	.9	2.07	2.82 (.002)	0	-
Ethnicity-Mixed	.98 (1.02)	13	.50	1.09 (.31)	.62	1.89	1.36 (.42)	.75	2.47		1.1 (.29)	.7	1.74	2.72 (.002)	0	-
Ethnicity-Other	2 (2.11)	26	5.74	.48 (.26)	.16	1.41	.7 (.39)	.24	2.1	**Reference group**	1.1 (.37)	.56	2.13	2.14 (.002)	0	-
Age	.29* (.16)	09	88	.7 * (.13)	.49	1	.62* (.13)	.4	.93		.65** (.1)	.49	.87	.62 (.44)	.15	2.51
FSM(Yes)	1.25 (.52)	55	.84	1.03 (.15)	.78	1.36	1.39* (.21)	1.03	1.87		1.2 (.14)	.95	1.5	1.49 (.74)	.56	3.96
SEN(Yes)	2.65^ (1.37)	96	.31	1.63* (.33)	1.1	2.42	1.85** (.4)	1.21	2.83		1.33(.23)	.94	1.88	2.67 (1.64)	.80	8.91
Academic Attainment	.88* (.05)	80	98	.93*** (.02)	.9	.97	.91*** (.02)	.87	.94		.96** (.01)	.93	.99	.93 (.06)	.82	1.06

***p< .001, **p< .01, *p< .05, ^p< .10

Table 4.5. Multi-level models predicting academic attainment (KS2) post the three waves

Parameter Estimates (Outcome: Key Stage 2)	**Model 1** *(baseline model)*	**Model 2** *(Model 1+ aggregate symptom score)*	**Model 3** *(Model 1+trajectories)*
	Estimate (SE)	Estimate (SE)	Estimate (SE)
Fixed Effects			
Intercept	21.55*** (1.36)	22.29*** (1.37)	22*** (1.37)
Prior Attainment: Key Stage 1	.89*** (.02)	.88*** (.02)	.88*** (.02)
Gender (Female)	-.32*** (.09)	-.27** (.09)	-.29** (.09)
FSM (Yes)	-.60*** (.13)	-.58*** (.13)	-.58*** (.13)
SEN (Yes)	-1.11*** (.18)	-1.07*** (.18)	-1.09*** (.18)
Ethnicity (Asian)	.69*** (.20)	.69*** (.20)	.69*** (.20)
Ethnicity (Black)	.13 (.27)	.15 (.27)	.14(.27)
Ethnicity (Mixed)	.01 (.25)	.01 (.25)	.01 (.25)
Ethnicity (Other)	1.20** (.39)	1.18** (.39)	1.18***(.39)
Relative age	-.80*** (.16)	-.84 *** (.16)	-.83*** (.16)
Aggregate internalising symptom scores		-.06*** (.02)	-
T1 *Decreasing v.high-moderate*			-.20 (.44)
T2 *Decreasing high-low*			-.28* (.13)
T3 *Stable high*			-.22 (.16)
T5 *Stable moderate*			-.35**(.11)
T6 *Increasing moderate-high*			-.19 (.54)
Variance Components			
Residual variance	2.54 (.03)	2.54 (.03)	2.54 (.03)
School-level	1.62 (.12)	1.62 (.12)	1.62 (.12)

***p< .001, **p< .01, *p< .05, ^p< .10

Results: 11-14 years

Descriptive statistics

The means, standard deviations and correlations of all the study variables are presented in Table 4.6.

Stage 1: Identifying heterogeneous developmental trajectories

Two- to seven-trajectory models were explored in each sample. Table 4.7 presents fit indices and criteria for model selection for the solutions obtained. In the secondary school sample the entropy of most of the models was similar and sufficient (>.70) and hence the decision was based on other indicators. The model with 5-classes was chosen for further exploration as the log-likelihood value clearly dropped between the 5-class and 6-class model and the LMR-LRT indicated that the 6-class model was not better fit to the data than the 5-class model. Additionally, the 6-class model did not identify any groupings that were distinct from the 5-class model and 5-class model offered significantly higher heterogeneity and posterior probabilities than the 4-class model (largest class 57% vs. 73%).

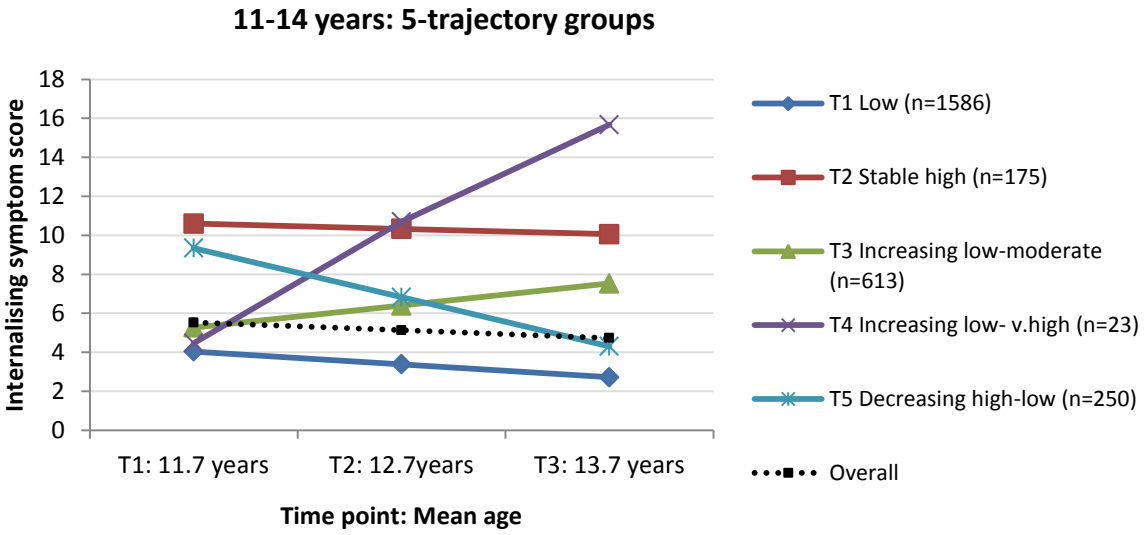


Figure 4.3. Developmental trajectories of internalising symptoms of adolescents aged 11-14 years

Table 4.8 presents the sample details and the intercept and slope co-efficients for the developmental trajectories in the selected 5-trajectory model and for the overall sample (single trajectory model).

As can be seen from Table 4.8 and Figure 4.3, in the older sample there were two trajectories of significantly increasing symptoms over time (23.2% and .9%) which indicates that almost a quarter of students significantly developed symptoms between ages 11 to 15. 9.4% (N=250) of the sample had significantly decreasing symptoms from high and a large majority (59%) had low decreasing internalising symptoms over the three waves.

Amount of school level variation in trajectories was small, ICC=.01, which resulted in design effect sizes (1+ (average cluster size-1)*ICC; (Muthén & Satorra, 1995) of less than 2 (1.73). Hence nesting was not accounted for in Stage 2 analysis (ibid.).

Stage 2: Correlates of heterogeneous symptom trajectories

Table 4.9 presents the MLR results for the secondary school 5- trajectory model. As can be seen in Table 4.9, in the older sample aged 11-15 years, being female significantly increased risk of being in all the higher symptom trajectories, except the increasing low-high trajectory where they showed significantly decreased risk. Females were more than 3 times as likely to have stable high symptoms compared to males (RR=3.43). Asian ethnicity decreased probability of being in the stable high group (RR = .63, $p<.05$). FSM increased risk of having an increasing low-moderate symptoms trajectory (RR = 1.28, $p<.10$) and SEN significantly predicted higher risk of having stable high and an increasing trajectory. Academic attainment significantly predicted the stable high and decreasing trajectories, with higher scores decreasing risk of having stable high symptoms (RR = .68, $p<.01$) and decreasing probability of having decreasing symptoms (RR = .79, $p<.05$).

Stage 3: Predicting change in academic attainment

Table 4.10 shows the results of the MLMs predicting Key stage 3 results in the secondary school sample, controlling for prior attainment (Key Stage 2). Model 1 includes prior attainment and all the co-variables. Model 2 includes aggregate symptom scores as a predictor and Model 3 includes the trajectories of emotional symptoms as categorical predictors of post-attainment, with the reference category being the low decreasing trajectory (Trajectory 1).

In the older cohort (see Table 4.10), similar to the younger cohort, Model 2 indicates that aggregate symptoms significantly predict change in attainment scores, with higher problems indicating negative gain in attainment. The co-efficient of change is low (β = -.01, p<.001, ES=.01). Looking at the model in which different trajectories predict scores (model 3), when compared to the low symptom trajectory pupils with increasing trajectories had significant or nearly significant negative co-efficients of change in attainment (increasing low-moderate β = -.05, p<.05, ES=.07; increasing low-high β = -.18, p<.10, ES=.20). The remaining two trajectory groups, stable high and decreasing symptom groups did not significantly predict worse change in attainment compared to the reference group (ES=.06 and .07 respectively).

Table 4.6. Means, standard deviations and correlations for variables in the 8-11 year old sample

	Mean (SD)/ Percentage	**Correlations**							
		1.	2.	3.	4.	5.	6.	7.	8.
1.T1Internalising	5.53(3.34)								
2.T2Internalising	4.84(3.25)	.52**							
3.T3Internalising	4.74(3.38)	.42**	.55**						
4.Pre-attainment	4.16(.66)	-.15**	-.08**	-.07**					
5.Post-attainment	5.54(.96)	-.15**	-.08**	-.09**	.81**				
6.Age at T1	11.71(.29)	-.06*	-.00	-.00	.08**	.06**			
7.Gender (Female)	54.4%	*.13***	*.15***	*.16***	*.00*	*.06**	*.01*		
8.FSM	17.1%	*.02*	*.05**	*.05**	*-.17***	*-.18***	*.01*	*.03*	
9.SEN	7.1%	*.09***	*.07***	*.08***	*-.27***	*-.26***	*.00*	*-.10***	*.06***

Note. Italicised are non-parametric correlations (Spearman's rho). *p<.10, **p<.001

Table 4.7. Fit indices and criteria for model selection for 2-7 class LCGA solutions

	BLRT	**LMR_LRT**	**A-BIC**	**Entropy**	**Posterior probabilities**	**% in classes (estimated)**
2- Class	188.57**	180.92**	39663.75	.68	.81-.92	20.8 - 79.2
3-Class	131.49 **	126.16 *	39546.37	.74	.77 -.93	6.6-77.1
4- Class	74.71 **	71.68*	39485.77	.74	.69-.90	6.2 – 72.7
5-Class	123.45**	118.44**	39376.43	.73	.72-.88	.9-57
6-Class	27.26**	26.15	39363.28	.75	.7-.88	.8- 55.3
7-Class	25.45**	24.42	39351.94	.74	.69-.87	.6-50

*p<.10, **p<.001

Table 4.8. Sample breakdown, socio-demographic descriptives and intercept and slope co-efficients for the trajectory groups in the secondary school sample (11-14 years)

Trajectory group	**N (%)**	**Gender % (Female)**	**FSM % (Yes)**	**Age M (SD)**	**SEN % (Yes)**	**Attainment M (SD)**	**Intercept**	**Slope**
T1 *Decreasing low*	1586 (59.9)	48.9	14.9	11.71 (.29)	5.5	4.20 (.64)	4.04	-.66**
T2 *Stable high*	175 (6.6)	74.9	20.7	11.71 (.29)	13.1	3.98 (.78)	10.6	-.27
T3 *Increasing low-moderate*	613 (23.2)	62.6	19.9	11.71 (.30)	7.8	4.14 (.64)	5.27	1.13**
T4 *Increasing low-high*	23 (.9)	26.1	17.4	11.67 (.26)	13	4.23 (.79)	4.47	5.6**
T5 *Decreasing high-low*	250 (9.4)	57.2	20.9	11.69 (.30)	10.4	4.05 (.70)	9.34	-2.52**
Overall sample	2637	54.4	17.1	11.71 (.29)	7.1	4.16 (.66)	5.53	-.4**

**<.001

Table 4.9. Relative Risk Ratios (RR) for the Multinomial logistic regression in the 5-trajectories in the secondary school sample (11-14 years)

Predictors		**Trajectory 1** *Decreasing low*	**Trajectory 2** *Stable high*		**Trajectory 3** *Increasing low- moderate*			**Trajectory 4** *Increasing low-high*			**Trajectory 5** *Decreasing high-low*		
			RR (SE)	**95% CI**		**RR (SE)**	**95% CI**		**RR (SE)**	**95% CI**		**RR (SE)**	**95% CI**	
Gender (Female)	Reference group		3.43*** (.66)	2.35	5.01	1.79***(.18)	1.46	2.18	.36* (.18)	.14	.94	1.51** (.22)	1.13	2.00
Ethnicity-Asian			.63* (.15)	.39	1.02	.88 (.12)	.67	1.14	2.02 (1.09)	.70	5.82	1.07 (.19)	.75	1.52
Ethnicity-Black			1.55 (.48)	.84	2.83	.95 (.22)	.60	1.50	1.72 (1.83)	.22	13.77	.92 (.31)	.47	1.78
Ethnicity-Mixed			1.27 (.58)	.52	3.10	1.10 (.32)	.63	1.93	4.52^ (3.55)	.97	21.04	.87 (.39)	.36	2.10
Ethnicity-Other			1.25 (.97)	.28	5.68	1.49 (.63)	.65	3.42	.00 (.004)	0	-	.00 (.001)	0	-
Age			.97 (.28)	.56	1.71	.96 (.16)	.69	1.34	.53 (.40)	.12	2.13	.82 (.20)	.51	1.31
FSM (Yes)			1.27 (.27)	.83	1.92	1.28^ (.17)	.99	1.66	.94 (.60)	.27	3.31	1.34 (.25)	.94	1.91
SEN (Yes)			2.06** (.61)	1.15	3.69	1.54* (.32)	1.03	2.31	2.43 (1.96)	.82	12.58	1.65^ (.44)	.98	2.77
Academic Attainment			.68** (.09)	.53	.87	.94 (.08)	.80	1.10	1.75 (.70)	.79	3.85	.79* (.09)	.64	.98

***p< .001, **p< .01, *p< .05, ^p< .10

Table 4.10. Multi-level models predicting change in academic attainment

Parameter Estimates (Outcome: Key Stage 3)	**Model 1** *(baseline model)* Estimate (SE)	**Model 2** *(Model 1+ aggregate symptom score)* Estimate (SE)	**Model 3** *(Model 1 + trajectories)* Estimate (SE)
Fixed Effects			
Intercept	1.15** (.41)	1.26** (.41)	1.20** (.41)
Prior Attainment: Key Stage 2	1.12*** (.02)	1.12*** (.02)	1.12*** (.02)
Gender (Female)	.06** (.02)	.07*** (.02)	.07** (.02)
FSM (Yes)	-.13*** (.03)	-.12*** (.03)	-.12*** (.03)
SEN (yes)	-.17*** (.04)	-.16*** (.04)	-.16*** (.04)
Ethnicity (Asian)	.04 (.04)	.03 (.04)	.03 (.04)
Ethnicity (Black)	.06 (.06)	.06 (.06)	.06 (.05)
Ethnicity (Mixed)	.06 (.06)	.06 (.06)	.06 (.06)
Ethnicity (Other)	.21* (.10)	.20* (.10)	.20* (.10)
Relative Age	-.02 (.03)	-.03 (.03)	-.03 (.04)
Aggregate internalising symptoms		-.01*** (.00)	-
T2 *Stable high*			-.06 (.04)
T3 *Increasing low-moderate*			-.05*(.025)
T4 *Increasing low-high*			-.18 ^(.11)
T5 *Decreasing high-low*			-.05 (.04)
Variance Components			
Residual variance	.50 (.01)	.50 (.01)	.50 (.01)
School-level	.25 (.04)	.25 (.04)	.25 (.04)

***p< .001, **p< .01, *p< .05, ^p< .10

Discussion

The current study aimed to utilise a trajectory based approach to better understand the association between development of internalising symptoms and changes in the academic attainment of primary and secondary school children. Analyses were conducted in multiple stages to first identify empirically derived trajectories of symptoms, assess the predictors that are associated with them and subsequently use the derived trajectories in a model predicting subsequent attainment after controlling for prior academic attainment.

Heterogeneous trajectories

The analyses indicate that, as hypothesised, heterogeneous growth trajectories better represent development of child and adolescent internalising symptoms over time. The single trajectory approach would have resulted in an overall decreasing trend over time. Instead, based on explorations of multi-trajectory models in the younger and older samples 6-trajectory and 5-trajectory models were identified respectively for further exploration. The identification of independent growth trajectories that better represent the data demonstrates the relevance of understanding heterogeneity in the development of psychopathology.

As would be expected from known prevalence of disorder in the community (Green et al., 2005) and studies of symptom development (Dekker et al., 2007), the largest proportions of individuals were in the group with low levels of symptoms at all time points (primary 40%, secondary 59%). In both age groups identified developmental trajectories included stable high symptoms, decreasing symptoms and increasing symptoms over three waves and additionally stable moderate symptom trajectory was identified in the younger sample. Proportion in the stable high groups in the primary sample (11%), were slightly higher than would be expected from known community levels of emotional problems, whereas in the older sample (6.6%) were similar to expected levels of problems in an English sample of this age range (Green et al., 2005). Of particular clinical relevance are the decreasing and increasing trajectories as they represent recovery and development of problems respectively. 17.5% children and 9.4%

adolescents had significantly decreasing symptom trajectories from initial clinical level scores. On the other hand, the younger sample identified a very small proportion of individuals with increasing symptoms from age 8-11 (.7%), whereas in the older sample more than 600 (24.1%) individuals had significant increases in symptoms from age 11-14, which supports epidemiological findings that emotional problems peak in adolescence around ages 14-15 (Costello et al., 2011).

Correlates of heterogeneous trajectories

Analyses indicate that socio-demographic variables predicted different developmental trajectories of symptoms to a certain extent. Gender was a key predictor of group membership and females were more likely to have stable moderate and high symptoms when compared to males in the younger sample. In the older sample, aged 11–14 years, being female predicted higher levels of symptoms overall. However, males demonstrated increased risk of having a steep increasing symptom trajectory in this age group. The finding of gender differences in the younger sample is contrary to studies that have found no gender differences in pre-adolescence (e.g., Wade et al., 2002) and indicates that gender differences in internalising symptoms are observed even before adolescence. This finding challenges theories of emotional symptoms that focus on puberty as the main trigger for gender differences in emotional symptoms (Nolen-Hoeksema & Girgus, 1994) and support the notion that gender differences exist before adolescence and are potentially widened by processes that occur during and after puberty (Hammen & Watkins, 2008).

Deprivation significantly increased probability of membership in the stable high trajectory in the younger group and the increasing low-high trajectory for the older group. This finding is supported by theories of financial stress predicting emotional stress (Baum et al., 1999) and links between socio-economic status and emotional symptoms that have been found in epidemiological studies (Green et al., 2005). Links between ethnic groupings and development of emotional symptoms in British samples have not been clearly established, one

reason being that they have not been studied as much (Nikapota & Rutter, 2009). The findings of this study indicate that belonging to the Asian ethnic category significantly increases risk of having stable moderate symptoms in pre-adolescence and decreases risk of having stable high problems in adolescence. Belonging to the Black ethnic group significantly increased the probability of being in the decreasing high-low group in the primary sample. Though these findings begin to give clues on socio-ethnic influences on the development of trajectories this area needs more in-depth explorations (Nikapota & Rutter, 2009) and investigation looking at sub-categories of the broad ethnic categories might help further disentangle these effects (e.g., Goodman et al., 2008). The findings of this study indicate that both deprivation and ethnic background are predictors of different developmental patterns in internalising symptoms, which is contrary to results from a large study with an American sample which found no effects of these factors on the persistence of symptoms in adolescents (Rushton et al., 2002). This difference could potentially be explained by methodological differences as breaking symptoms down to heterogeneous groups allows more sensitive analysis of sub-groups of the sample, or on the other hand, might represent geographical variation in the relevance of these variables; the former being more likely as cross-sectional studies in American samples have found deprivation to be a strong predictor of emotional disorder (e.g., Gilman et al., 2002).

A unique contribution of this study in terms of correlates was the inclusion of school-related risk factors: SEN and relative age, that might be associated with symptom development. SEN significantly increased the risk of having stable high symptoms and decreasing symptoms in the younger sample; and both increasing decreasing symptoms in the older sample. SEN predicting higher symptoms would be expected (Humphrey et al., 2013), the inconsistency in the direction of SEN predicting the development of symptoms might be due to various factors including a) various sub-categories of SEN (needs could be cognitive, social , learning, speech, sensory or physical; for instance dyslexia, autism spectrum and physical disabilities all count as SEN), b) different levels of help and support received, or c)

some, but not all individuals with SEN developing resilience based on personal experiences (Patterson & Blum, 1996). Hence factors such as resilience, personal experience, coping mechanisms and different types of SEN are some of the factors that are associated with having SEN that might contribute differently to symptom development in these children (Lyon, 1996; Masten, 2001).

In the younger sample another steady predictor of trajectories was relative age, with younger pupils within the cohort being significantly more likely to belong to most high symptom groups. This is an effect that has previously been found in a large cross-sectional study of mental health (Goodman et al., 2003) but has not been explored in the longitudinal context for mental health. However, the negative effects of relative age have been documented in other child developmental domains such as self-esteem (Thompson et al., 2004) and academic performance (Boardman, 2006; Fredriksson & Öckert, 2005). In the older sample this effect was not found which suggests that the effect of developmental lag experienced by younger pupils in primary schools potentially weakens by the time they are in secondary school and hence does not directly predict higher levels of problems. The size and consistency of the relative age effect in pre-adolescence also suggests that it should be included in clinical practice and case consideration to help identify individuals at increased risk of emotional disorder. This is a predictor that would benefit more detailed exploration in relation to long term developmental effects on mental health disorders.

Predicting change in academic attainment

The final stage of analysis attempted examine the association between internalising symptom development and academic attainment, alongside comparing the utility of using aggregated symptoms versus developmental trajectories to predict outcomes. Past analysis using aggregated symptoms to predict change in academic attainment (e.g., Cole et al., 1996) found no link of symptoms on relative gains in learning.

The results of this stage of analysis indicate that developmental trajectories differentially predict changes in academic attainment over the same period of time. In the younger primary school sample having moderate stable symptoms had an adverse impact on academic attainment after controlling for all other socio-demographic predictors. For this age group, having a decreasing symptom trajectory also negatively predicted academic attainment. These results have serious developmental and clinical implications as they demonstrate that: 1) sub-clinical levels of symptoms have adverse effects on other areas of child functioning, and 2) even when symptoms decrease there is a persistence of the developmental lag caused by high symptoms at an earlier stage in development, leading to reduced academic performance even when the symptoms have subsided.

In the secondary school sample, increasing symptom trajectories over the three years adversely impacted on academic performance post wave three, which indicates that increasing problems were associated with a significant decrease in performance while controlling for previous attainment. This is consistent with the hypothesis that difficulties in one area of functioning, such as mental health, can have knock-on effects on other domains of functioning.

Additionally, when compared to utilising the aggregated symptom score as a predictor of academic attainment, the trajectories present a more nuanced picture in which different trajectories predict change in academic attainment to varying degrees. It is important to note here that even when using an aggregated approach in this study the results indicate an association between level of symptoms in the three years and change in academic attainment. This contrary finding to some previous research (Cole, 1990) might be to some extent due to the different types of attainment measures used, with the previous studies having used child or proxy evaluations of academic competence instead of standardised national assessments.

As outlined in the introduction to the chapter, negative development of internalising symptoms might be expected to predict negative development in the same and other associated domains, such as attainment (Masten et al., 2005). The results of this study support this general notion, with the secondary school results clearly indicating that increasing internalising symptoms were associated negatively with subsequent attainment. However, the primary school results also suggest that deficits at one stage in development, although followed by positive development (decrease in symptoms) can still adversely affect development in other domains. We postulate that this occurs due to the impact of deficits in one domain in early development (e.g., high internalising symptoms) on functioning in other domains (e.g., academic attainment) when they occur and even though there is subsequent positive development in one domain it is more difficult for the child to 'catch-up' in the other domain and as a result experience a lag in development (e.g., lower academic attainment). The occurrence of this effect in the younger sample might also suggest that these effects might be more salient at earlier developmental stages when hindrances to learning have more serious long-term implications. This finding would benefit from more exploration across different samples and domains to assess whether it is replicated.

The current analyses shed some light on the longitudinal impacts of internalising symptoms on change in academic attainment and results indicate that developmental trajectories present a more nuanced picture where different trajectories predict change in academic attainment to varying degrees. In the primary sample having moderate stable symptoms has an adverse impact on academic attainment after controlling for all other socio-demographic predictors. Belonging to the decreasing trajectory also negatively predicted academic attainment. These results have serious developmental and clinical implications as they demonstrate that 1) sub-clinical levels of symptoms have adverse effects on other areas of child functioning and 2) even when symptoms decrease there is a persistence of the developmental lag that high symptoms at a prior stage caused, leading to reduced academic

performance even when the symptoms have subsided. On the other hand, in the adolescent sample increasing symptom trajectories in the three waves adversely affected academic performance post the wave three, which indicates that increasing problems was associated with a significant decrease in performance while controlling for previous attainment. This is in line with the hypothesis that difficulties in one area of functioning such as mental health can have knock on effects on other domains of functioning.

In summary the results indicate that studying the risk factors and predictive associations of heterogeneous developmental trajectories of internalising symptoms allows a more dis-aggregated understanding both of the predictors of development of emotional problems and the academic risk associated with different developmental trajectories. The results clarify and in some cases further existing knowledge by using newer methodological approaches to revisit these issues.

Strengths and limitations

Similar to the previous study, the strengths and limitations of this study include two large cohorts of different age groups and non-random attrition respectively (see pages 158-159 in the previous study for detailed discussion).

A methodological drawback in deriving trajectories as shown in this study is the small size of some trajectory groupings. While these trajectory groupings are clinically meaningful and important to identify, when they are used as predictors of outcomes their lack of sufficient power must be considered when comparing the results of different groups. Suggested solutions for this power issue include using larger samples to help achieve higher numbers in trajectories with lower proportions and/or focussed studies with higher proportions of at-risk individuals (Dekker et al., 2007).

Implications and future directions

Over the past few years the focus of educational reforms has been on curriculum goals that are more academic and skills oriented resulting in social and emotional components of education taking a back seat (Shoshani & Steinmetz, 2013). The results of the current study lend powerful support to the arguments for prevention, early intervention and school-based support for mental health difficulties, and the need for greater integration between prevention and educational policy (Greenberg, 2010). The finding that stable symptoms even when sub-threshold have negative impact on later attainment support the need for universal approaches that focus on prevention of problems and promotion of well-being alongside reactive approaches after problems have arisen (Shoshani & Steinmetz, 2013).

In summary, a trajectory-based analysis provides an additional perspective in the interpretation of risk factors as potential indicators of the developmental path of a child rather than simply predicting the severity of emotional disorder. The results of the present study contribute to understanding which individual factors increase the risk of developing and

sustaining problems. Aside from their importance in terms of public health, they can be used to guide clinical practice, especially in terms of predicting the course, duration, and complexity of treatment, which also feeds into estimations of resource allocation. In terms of the longitudinal developmental ramifications of internalising symptoms on educational outcomes, the study adds to existing knowledge of these relationships by permitting a breakdown of the differential effects of different developmental symptom trajectories on academic attainment.

The results of this study indicate that using person-centred heterogeneous developmental trajectories to summarize differential development of internalising symptoms allows us to better estimate the longitudinal associations between internalising symptoms and academic attainment. The study provides a more definitive answer to questions regarding longitudinal impacts of internalising symptom development in children and at the same time provides a more nuanced picture of the academic risk associated with differential symptom development.

These two studies, this and the previous one, identified short-term developmental trajectories of internalising and externalising symptoms in the same samples in two cohorts of 8-11 years and 11-14 years. The existence of heterogeneous symptom trajectories that individuals can be grouped into based on their symptom development, leads to questions of possible patterns in the simultaneous symptom development in the two domains. Existing research exploring issues of co-morbidity in development of symptoms have found that the symptom development slopes of these two domains correlate .4-.6 (Keiley, Bates, Dodge, & Pettit, 2000; Measelle, Stice, & Hogansen, 2006). The associations between symptoms in one domain predicting symptoms in the other domain are not yet conclusive, with studies suggesting early externalising symptoms predict later internalising, although a systematic review concluded that as there are not enough studies yet exploring this relationship only tentative conclusions could be drawn (Cerdá et al., 2008). There is also evidence for

internalising symptoms predicting change in externalising symptoms, for instance (Beyers & Loeber, 2003) found that in adolescent boys internalising symptoms predicted externalising symptoms more than externalising predicted internalising. The co-development of symptoms in these two key childhood psychopathology domains requires more exploration and is the topic of the next study.

Study 5. Co-development of internalising and externalising symptoms: a brief investigation

Background

Co-morbidity

Co-morbidity in psychopathology used to be defined based on concurrent co-occurrence of symptoms as the "co-occurrence of two or more separate child psychiatric conditions" (Caron & Rutter, 1991). However, the definition has since incorporated sequential co-morbidity: the presence in sequence of two or more disorders in an individual (Angold, Costello, & Erkanli, 1999; Cerdá et al., 2008), studies of which are possible in longitudinal datasets. Angold et al. (1999) also make the distinction between homotypic and heterotypic co-morbidity, where homotypic indicates co-morbidity between disorders within a grouping (e.g., depression and anxiety within the grouping of internalising) and heterotypic refers to co-morbidity of disorders from different diagnostic grouping or dimensions (e.g., depression and conduct disorder).

Co-morbidity in psychopathology is associated with, among other things, higher functional disability, longer durations of illness, lower quality of life and decreased social competence (Kauer-Sant'Anna et al., 2007; Renouf, Kovacs, & Mukerji, 1997). Additionally, co-morbidity is associated with higher service use (Du Fort, Newman, & Bland, 1993).

Proposed explanations of co-morbidity

Several hypotheses and theories have been suggested to explain co-morbidity of symptoms in psychopathology (Caron & Rutter, 1991; Lilienfeld, 2003). The proposed reasons broadly encompass methodological factors or substantive factors (Lilienfeld, 2003). Methodological factors are linked to errors or biases in measurement and include 1) shared diagnostic criteria, where overlap is due mainly to similar symptomology, 2) referral biases, which include two main types: Berkson bias (Berkson, 1946), which can be summarised as "mathematical consequence of the fact that an individual with two disorders can obtain treatment for either disorder" (Lilenfeld, 2003, p.286), and clinical selection bias; where

individuals with more than one disorder more likely to seek treatment as they experience greater impairment (Du Fort et al., 1993). The substantive reasons that are offered in the literature are briefly summarised below.

Shared risk/aetiology. A frequent hypothesis is that shared risk factors might explain co-occurrence of disorder types as disorders are multi-factorial in origin and many causal/risk factors of disorders are not disorder specific (for e.g., deprivation, conflicted family life). This mechanism can also be understood as *multifinality*, where a specific adversity or event might lead to different psychopathological outcome in different individuals (Cicchetti & Rogosch, 1996). This shared risk also possibly represents a combination of genetic and environmental risk (Silberg et al., 2003), as has been observed in twin studies of psychopathology (Lahey, Van Hulle, Singh, Waldman, & Rathouz, 2011).

Co-occurring risk. Another hypothesis suggest that co-morbidity might result from an overlap of risk factors; where disorders have distinct risk factors co-morbidity might occur when the risk factors tend to co-occur leading to the manifestation of different symptoms or disorders. Although this hypothesis is theoretically plausible, it does not have empirical support as yet, mainly due to the limited number of disorder specific risk factors that can be identified. In most investigations so far the risk factors for several different disorders seem to be more similar than distinct (for e.g., child maltreatment in a risk factor for different disorders including internalising and externalising (Green et al., 2010).

Common syndrome. Caron and Rutter (1991) posit the hypothesis that the pattern of co-morbidity might in itself constitute a meaningful syndrome. This hypothesis has some support especially in child psychopathology for instance where depression and anxiety are not highly distinct in children, their co-morbidity is explained by them both being internalising syndromes with a lot of overlap in symptomology. There is also research suggesting that there is a higher likelihood of these disorders being misclassified by practitioners in child

psychopathology precisely for reasons of overlap in symptoms. However, this explanation does not completely support heterotypic co-morbidity between symptoms belonging to different domains such as, for instance conduct from externalising and depression from internalising.

Pathogenetic comorbidity. Where one disorder predisposes or creates an increased risk for another (Lilienfeld, 2003). The primary disorders might indirectly increase risk for the subsequent disorder by influencing neurophysiological, individual or social changes that increase vulnerability for acquiring subsequent disorders or the primary could directly cause the onset of the subsequent disorder (Cerdá et al., 2008). Studies of heterotypic and homotypic sequential co-morbidity in psychopathology focus on this explanation and many theories have evolved for the longitudinal predisposition to experience a certain disorder based on already or previously having symptoms of another. Examples for the specific theories for explaining the longitudinal relationships between internalising and externalising are discussed in the next section.

Co-morbidity of externalising and internalising symptoms in young people

Epidemiologic studies of cross-sectional samples have reported on moderate associations between internalising and externalising symptoms, ranging up to 0.6 (Achenbach & Rescorla, 2001; Measelle et al., 2006). In terms of percentage of overlap, for instance (Garnefski & Diekstra, 1997) report that in a large community based sample of adolescence 29-38 % who have symptoms of internalising disorders also have symptoms of externalising disorders at the same time.

In terms of longitudinal co-occurrence and co-development of symptoms, there is a dearth of studies that have focussed on co-development, with most studies focussing on sequential co-morbidity. Measelle et al. (2006), examined associations in the development of internalising and externalising symptoms in their sample of adolescent females (n=493, age

12-15 years) and based on a latent growth curve approach examined correlations at baseline (r=.57) and correlations in slopes (r= .41). This indicates that, although internalising and externalising have distinct development patterns they are not only correlated cross-sectionally but also correlated moderately in development. They also found that initial externalising symptoms predicted increasing internalising symptoms during adolescence, and that initial internalising symptoms predicted a slower deceleration of externalising symptoms. Using a similar approach, Keiley et al. (2000) investigated longitudinal associations between baseline scores and slope of mother and teacher reported internalising and externalising scores from kindergarten to seventh grade (n=405, age 5 – 12 years) and found that relationship between intercept and slope in each domain varied based on reporter (mother or teacher), mother reported internalising and externalising slopes were highly associated (r=.6) but teacher rated ones were not associated (r=.02). Another study focussing on concurrent development of symptoms in a sample of adolescent boys (n=506, 13.5 years at baseline) found that internalising and externalising were both associated with development of symptoms in the other domain concurrently; longitudinally they found that internalising predicted development of externalising more than externalising predicted development of internalising symptoms (Beyers & Loeber, 2003).

However, such studies of associations in symptom development are less common and have been limited to specific gender or age groups. Studies of their longitudinal associations have tended to focus on sequential co-morbidity of symptoms, trying for the large part to determine whether internalising symptoms lead to externalising or vice versa. For instance, externalising behaviours during childhood and adolescence have been shown to predict subsequent internalising symptoms in later adolescence and early adulthood (Capaldi, 1992; Kim, Capaldi, & Stoolmiller, 2003).

There are indications that to the greater extent externalising symptoms may precede internalising symptoms, however, Cerdá et al. (2008) conducted a meta-analysis and

concluded that due to the limited number of studies that have examined these relationships longitudinally, especially reciprocal relationships, it is yet too tentative to draw definitive conclusions.

With regards to the hypothesised mechanisms involved in the longitudinal relationships between internalising and externalising disorders, several alternate theories have been proposed. The 'failure model' which hypothesizes that externalising problems disrupt interpersonal functioning which could lead to more strained relationships with family, peers and others, social rejection and failure in domains such as academics. The theory hypothesises that persistent failures in adaptation then contribute to the onset and development of internalising symptoms (Capaldi, 1992; Capaldi & Stoolmiller, 1999). The converse theory posits that internalising symptoms might interfere with normal social and emotional development. Subsequently this reduces concerns about the consequences of externalising behaviours leading to lower functioning in inter-personal relationships, interpersonal conflict and increasing risk for externalising behaviours (Capaldi, 1991; Cerdá et al., 2008). Both theories have some empirical support and it is likely that both these and other mechanisms are all involved, variously contributing to symptom development. There is evidence that common factors might underlie the aetiology of both internalising and externalising dimensions of child psychopathology, for instance, O'Connor, McGuire, Reiss, Hetherington, and Plomin (1998) used twin data to demonstrate that 45% of the observed co-variation in internalising and externalising was explained by genetic liability. Other proposed factors that have been suggested to contribute to shared aetiology of internalising and externalising in childhood include emotional maladjustment (Lilienfeld, 2003; Watson & Clark, 1984), temperament (Clark, 2005; Krueger, Caspi, Moffitt, & Silva, 1998), and personality factors (De Pauw & Mervielde, 2010).

Existing literature into person-centered longitudinal trajectories of symptoms of internalising and externalising disorders have usually been quite distinct, with studies

focussing on one or the other domain. No study was identified where trajectories of internalising and externalising over the same period were examined in the same sample for co-occurrence in types of symptom trajectories. The benefits of this approach include a possibility of identifying concurrent symptom development patterns and the proportions of individuals that have possible combinations of con-current symptom development. The patterns and proportions might provide some insight into heterogeneity in developmental co-morbidity in late childhood and early adolescence.

The current study

The current study aimed to map out the person-centered trajectories in developing internalising and externalising symptoms derived from the previous two studies in this chapter to examine the concurrence in development of these two dimensions in the two cohorts (8-11 years and 11-14 years). To do this, individuals' memberships to the different trajectory groups for both internalising and externalising will be tabulated. Additionally, the study aims to determine baseline (intercept) and change (slopes) in both internalising and externalising symptoms using a latent curve approach to examine the co-development of symptoms over time in both children and early adolescents.

It is hypothesised that there will be a moderate degree of co-development of internalising and externalising symptom development in both age groups. It is also hypothesised that intercept in each domain will predict symptom development in the other domain, independent of the influence of the intercept of the same domain.

Method

Participants

The participants are the same as in studies 3&4 (see pages 119-121 for details). Briefly, the primary school sample consisted of a total of 3346 individuals aged an average age of 8.7 years wave 1 and the secondary school sample consists of 2647 students with an average age of 11.7 years at time 1.

Key Variables

Externalising symptom trajectory

Based on the person-centered trajectories in development of externalising symptoms that were derived previously in Study 3; individuals were assigned to one of 4 groupings based on their trajectory membership: low symptoms (coded 1), stable high-moderate symptoms (coded 2), decreasing symptoms (coded 3) and increasing symptoms (coded 4). In the primary school sample of 3346 individuals this resulted in 1644 with a low symptom trajectory, 96 with stable symptoms, 454 with a decreasing symptom trajectory and 1151 with an increasing symptom trajectory. In the secondary school sample from age 11-14, out of 2647 individuals, 1597 had a low symptom trajectory, 275 a stable symptom trajectory, 600 a decreasing symptom trajectory and 175 had an increasing symptom trajectory.

Internalising symptom trajectory

Based on the person-centered trajectories in development of internalising symptoms that were derived in Study 4 individuals were assigned to one of 4 groupings based on their trajectory membership: low symptoms (coded 1), stable symptoms (coded 2), decreasing symptoms (coded 3) and increasing symptoms (coded 4). Based on this grouping criterion this resulted in the following groupings in the primary school and secondary school samples. In the primary school sample from age 8-11 years 1365 participants had a low symptom trajectory, 1369 had stable moderate- high symptoms, 587 had decreasing symptoms and 25

had increasing symptoms. In the older sample out of 2647 participants, 1586 had low symptom trajectories, 175 had stable symptoms, 250 had decreasing symptom trajectories and 636 had an increasing internalising symptom trajectory.

Externalising symptoms

M&MS behavioural difficulties scale (see study 3 Methods for description of measure).

Internalising symptoms

M&MS emotional difficulties scale (see study 4 Methods for description of measure).

Analytic Strategy

To investigate the relationships in the development of internalising and externalising symptoms over time two analyses were carried out to examine the associations or co-morbidity in symptoms and their development over time. The first analysis utilised the person-centered identified trajectories from the preceding two chapters and the second analysis created latent intercept and slope variables to allow examinations of relationships between baseline symptoms and change in symptoms.

Analysis 1: Trajectory membership for externalising and internalising from study 3 &4 respectively were mapped onto each other. Proportions were estimated based on both, proportion of individuals with a particular internalising trajectory who have each of the possible externalising developmental trajectories and vice-versa. Chi-square analyses were conducted first for the entire cross-tabulated trajectories table to determine whether the pattern overall were significantly different from chance proportions. Next, chi-square analysis was conducted to determine for every type of internalising or externalising trajectory, the corresponding trajectories in the other domain were distributed in a pattern significantly different from chance. For instance, for the low internalising trajectory group whether the

corresponding types of externalising symptom trajectories in those individuals were significantly predicted.

Analysis 2: The second analysis was carried out in two stages. In the first stage latent growth curve analysis (LGCA) was conducted with each of the constructs (internalising and externalising) to compute two latent dimensions each: an intercept, and a slope, representing initial levels and change over time respectively. LGCA's were first carried out for the internalising and externalising variables separately and then the latent intercept and slope factor scores were outputted and then these scores were used in following analysis to examine longitudinal associations of change in these two dimensions. Figure 5.1 pictographically demonstrates LGCA and how the model is set-up, which follows the standard methodology used for computing latent intercept and slope in LGCAs with longitudinal data (Curran & Hussong, 2003; Muthén & Khoo, 1998). The symptom scores for either externalising or internalising at each available time point are set to load equally on the intercept (in this case 1) and as the time points are evenly spaced they are set to load, in this case 0, 1, 2 on the slope whereby the slope value, like in the regression framework, represents change for every unit increase in time. Based on this model factor scores for each individual representing their intercept and slope can be outputted and can be utilised in further analysis.

In Stage 2 the associations between the latent variables were examined using correlations to determine the extent to which baseline scores and change in symptoms in internalising and externalising symptoms were associated. Subsequently, regression analyses were conducted to examine the extent to which intercept scores in each domain predicted slope in the same and other domain while controlling for the intercept of the other domain.

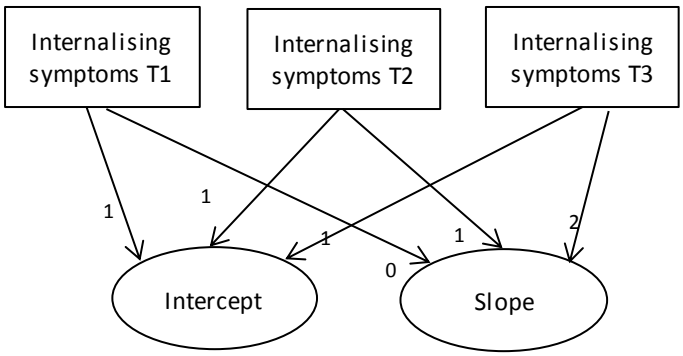


Figure 5.1. Latent growth curve model with manifest internalising symptoms at each time point and latent intercept and slope variables.

Results

Analysis 1

Tables 5.1 and 5.2 contain trajectory membership for internalising (rows) and trajectory membership for externalising (columns) in the 8-11 years and 11-14 years cohort samples respectively. The tables also include cells that indicate the proportions of a certain internalising trajectory having a certain externalising trajectory and vice versa. The last column and row present chi-square values that indicate whether the category membership in the other domain can be ascribed to just to chance or not. Figure 5.2 (a&b) represent the breakdown of other domain trajectory membership for each type of trajectory in the internalising and externalising respectively in the 8-11 years sample and Figure 5.3 (a&b) present the same information for symptom development from 11-14 years.

Table 5.1. Trajectory membership for internalising and externalising in the primary school sample (age 8-11years)

			Externalising Trajectory								
			(1) Low		(2) Stable high		(3) Decreasing		(4) Increasing		
			N	% in (1)	N	% in (2)	N	% in (3)	N	% in (4)	**χ^2 (df=3)**
Internalising Trajectory	(1)Low	N	894	54.4%	15	15.5%	132	29.1%	324	28.1%	
		% in (1)	65.5%		1.1%		9.7%		23.7%		**253.41*** **
	(2)Stable high	N	499	30.4%	57	58.8%	183	40.3%	630	54.7%	
		% in (2)	36.4%		4.2%		13.4%		46.0%		**179.65*** **
	(3)Decreasing	N	251	15.3%	18	18.6%	138	30.4%	180	15.6%	
		% in (3)	42.8%		3.1%		23.5%		30.7%		**60.69*** **
	(4)Increasing	N	0	0.0%	7	7.2%	1	0.2%	17	1.5%	
		% in (4)	0.0%		28.0%		4.0%		68.0%		**77.09*** **
	χ^2 (df=3)		**268.01*** **		**79.51*** **		**69.15*** **		**166.82*** **		

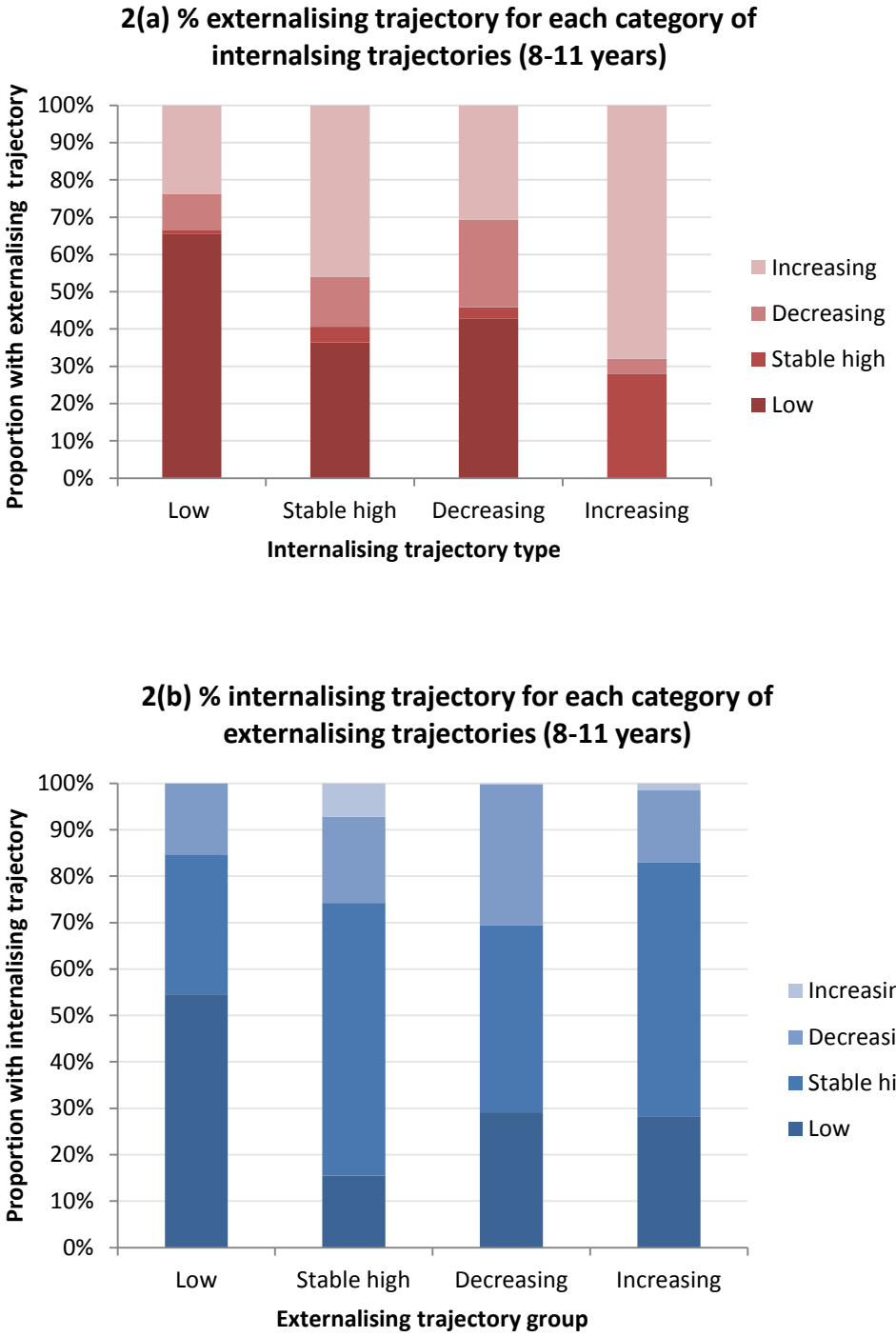


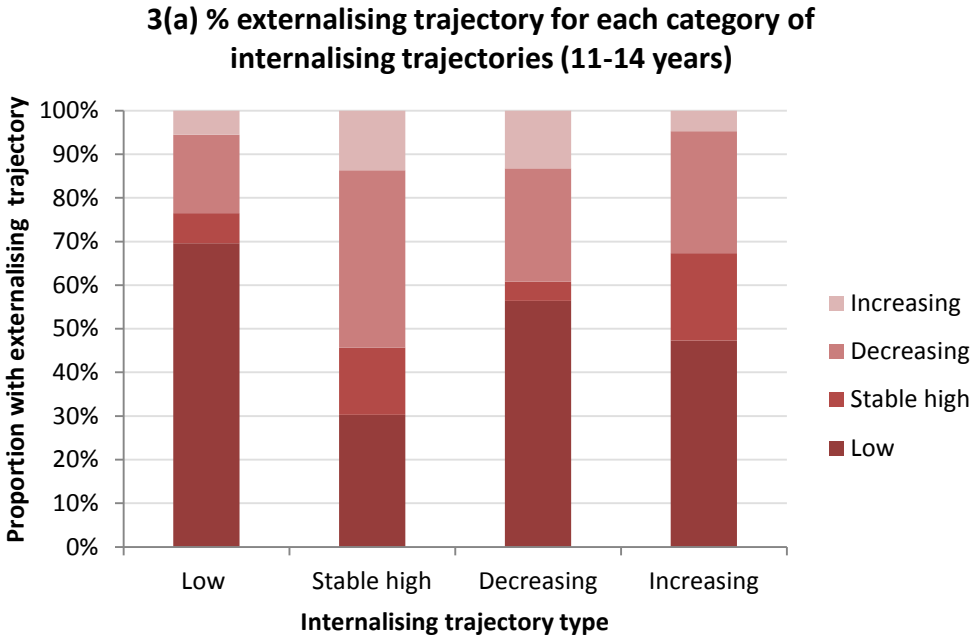
Figure 5.2. Figures showing the proportions of individuals with other domain trajectory types for internalising trajectories (2a) and externalising trajectories (2b) from 8-11 years

As can be seen in Table 5.1, in the younger sample more than half of individuals with low symptom trajectories in one domain had low symptom trajectories in the other (row 1 column 1) with no individuals with low externalising having increasing internalising

symptoms. In contrast around quarter of individuals who had low internalising trajectories had increasing externalising symptoms in the same time period (see row 1, column 4). The large majority (82.4%) of individuals with stable internalising symptoms had either low or increasing externalising symptoms over the three waves; whereas ~60% of individuals with stable externalising symptoms had stable internalising symptoms as well. 68% of individuals with increasing internalising had increasing externalising symptoms whereas the converse was true only of 1.5%. Overall, a greater number of individuals had increasing externalising rather than internalising symptoms in this age-group.

Table 5.2. Trajectory membership for internalising and externalising in the secondary school sample (age 11-14 years)

			Externalising Trajectory								
			(1) Low		(2) Stable high		(3) Decreasing		(4) Increasing		
			N	% in (1)	N	% in (2)	N	% in (3)	N	% in (4)	**χ^2 (df=3)**
Internalising Trajectory	(1)Low	N	1102	69.0%	110	40.0%	286	47.7%	88	50.3%	
		% in (1)	69.5%		6.9%		18.0%		5.5%		**144.59*** **
	(2)Stable high	N	53	3.3%	27	9.8%	71	11.8%	24	13.7%	
		% in (2)	30.3%		15.4%		40.6%		13.7%		**73.42*** **
	(3)Decreasing	N	141	8.8%	11	4.0%	65	10.8%	33	18.9%	
		% in (3)	56.4%		4.4%		26.0%		13.2%		**29.72*** **
	(4)Increasing	N	301	18.8%	127	46.2%	178	29.7%	30	17.1%	
		% in (4)	47.3%		20.0%		28.0%		4.7%		**112.41*** **
	χ^2 (df=3)		**168.07*** **		**97.41*** **		**63.29*** **		**38.48*** **		



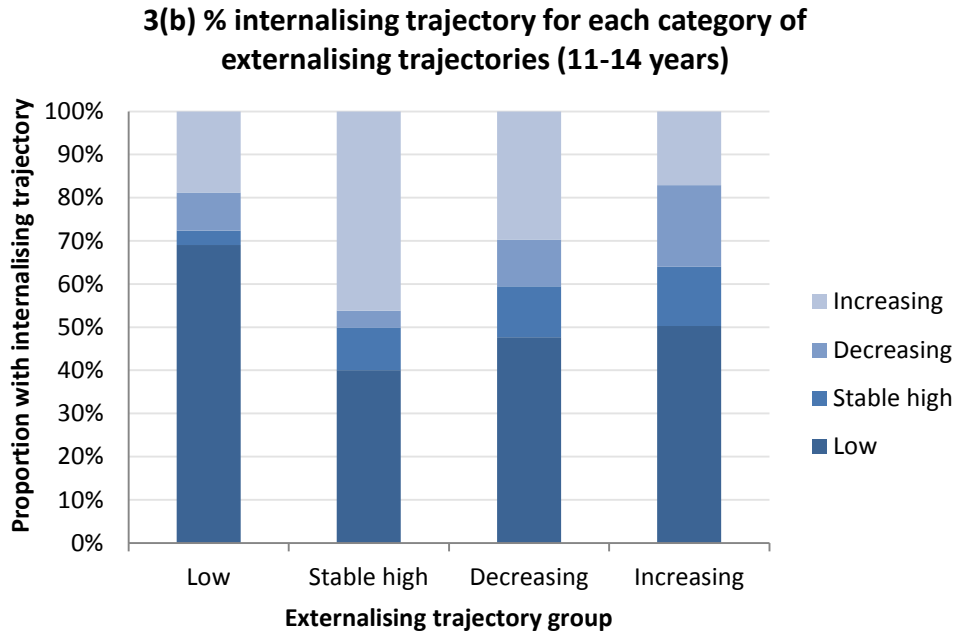


Figure 5.3. Figures showing the proportions of individuals with other domain trajectory types for internalising trajectories (3a) and externalising trajectories (3b) from 11-14 years

Results in Table 5.2 indicate that almost 70% of individuals with low trajectories in one domain had low trajectories in the other domain (row 1, column 1). The majority of individuals with stable internalising symptoms had either low (~30%) or decreasing

externalising symptoms (~40%), whereas ~46% of individuals with stable externalising symptoms experienced increasing internalising symptoms. Around half of individuals with increasing internalising or externalising symptoms had low trajectories in the other domain. In this age group almost a third of individuals with decreasing externalising symptoms experienced increasing internalising symptoms, whereas a similar proportion of individuals with decreasing internalising experienced decreasing externalising.

Analysis 2

Stage 1: LGCAs

LGCAs were estimated for both internalising and externalising symptoms over the three time points in both the primary and secondary schools samples.

Model fit criteria are presented in Table 5.3 below. In summary the model fit based on the CFI and TLI indicated very good model fit and based on RMSEA indicated very good fit for three of the four models, except internalising in the older sample. Means and variances for both the intercept and slope were estimated and were all statistically significant. However, looking at the values we can ascertain that, for instance, variance in internalising slopes were larger in the younger sample (variance=1.52) than the older sample (variance=0.99).

Table 5.3. Model fit statistics and means, variances of slopes and intercepts for each of the LGC models estimated

	CFI;TLI	RMSEA	Intercept		Slope	
		(90% CI)	mean	variance	mean	variance
8-11 years						
Internalising	.99;.99	.02 (.00-.06)	7.00***	6.85***	-0.74***	1.52***
Externalising	1; 1	.00 (.00-.04)	3.18***	3.95***	-0.17***	0.57***
11-14 years						
Internalising	.98;.95	.11 (.08-.14)	5.43***	6.36***	-0.39***	0.99***
Externalising	.99;.99	.04 (.01-.07)	3.14***	3.65***	-0.07**	0.41***

Stage 2: Relationships between derived intercepts and slopes

Intercept and slope factor scores were outputted from the four LGCA models in Stage 1 and used for the present analysis. Tables 5.4 and 5.5 present the correlations between the internalising and externalising intercept and slope values for the primary and secondary school samples respectively.

Table 5.4. Correlations between intercept and slopes of internalising and externalising symptoms in the primary school aged sample (8-11 years)

	Correlation co-efficients		
	Internalising intercept	Internalising slope	Externalising intercept
Internalising intercept	-		
Internalising slope	-.03	-	
Externalising intercept	.41***	.03	-
Externalising slope	-.05**	.37***	-.12***

Results in the younger sample in Table 5.4 indicate that, as might be expected from existing literature, the internalising and externalising intercepts are correlated moderately (r=.41). However, the correlations between internalising intercept and internalising slope (r=-.03, p=.12) and externalising intercept and internalising slope (r=.03, p=.07) were not statistically significant. The slopes for internalising and externalising symptoms were also correlated moderately (r=.37) indicating a positive association in development of symptoms in

these two domains. The slope of externalising symptoms was also correlated significantly and negatively with both the intercepts of internalising and externalising, indicating that higher intercepts were associated with lower externalising slope and vice versa.

Results for the correlations in the older sample can be seen below in Table 5. Intercepts of internalising and externalising (r=.34) and slopes of internalising and externalising (r=.36) were positively, significantly and moderately correlated.

Table 5.5. Correlations between intercept and slopes of internalising and externalising symptoms in the secondary school aged sample (11-14 years)

	Correlation co-efficients		
	Internalising intercept	Internalising slope	Externalising intercept
Internalising intercept	-		
Internalising slope	-.10***	-	
Externalising intercept	.34***	.02	-
Externalising slope	-.002	.36***	-.15***

In the case of the older sample (see Table 5.5), similar to the younger sample the intercept of externalising was not significantly associated with the slope of internalising (r=.02, p=.28), but in the case of the internalising the slope was negatively and significantly associated with the intercept of internalising (r= -.10). In the older sample the slope of externalising was not associated with the intercept of internalising (r=-.002, p=.92), but was significantly and negatively associated with the slope of externalising (r= -.15).

Table 5.6. Results of regression analysis predicting slopes of internalising and externalising symptoms in both age groups

	8-11 years		**11-14 years**	
Predictor	Internalising slope B(SE)	Externalising slope B(SE)	Internalising slope B(SE)	Externalising slope B(SE)
Intercept	-.64***(.045)	.07**(.025)	-.28***(.029)	.06**(.018)
Internalising intercept	-.02**(.006)	.005 (.003)	-.04*** (.004)	.02*** (.003)
Externalising intercept	.05***(.008)	-.04*** (.004)	.04***(.006)	-.04***(.01)
Internalising slope	---	.21***(.008)	---	.23***(.01)
Externalising slope	.70***(.029)	---	.61*** (.029)	---

Table 5.6 demonstrates the results of regression models predicting slopes in each domain including intercept of both domains and slope of the other domain. In the younger sample, results indicate that the intercept of externalising symptoms significantly predicted slope of internalising symptoms (B=.05, p<.001) while the intercept of internalising symptoms does not significantly predict development of externalising symptoms. In the older age group both the internalising and externalising intercepts significantly predicted slopes in the same and the other domain. The directionality of the relationship followed a similar pattern as well; same domain intercept predicted negative change in slope but other domain intercept predicted positive change in slope.

Discussion

The aim of the present analysis was to examine the co-morbidity of development of internalising and externalising symptoms over time. This was investigated by examining 1) the co-occurrence of different developmental trajectories of internalising and externalising symptoms over the same time period in late childhood and early adolescence and, 2) examining associations between latent intercept and slope scores across all individuals.

Co-occurring internalising and externalising trajectories

The first observation related to types of trajectories in internalising and externalising symptom development in the two age groups is the high proportions of children with increasing externalising trajectories and adolescents with increasing internalising trajectories. These proportions are in line with the established findings in existing literature where childhood is more dominated by externalising behaviours and adolescence by greater internalising symptoms (Martel, 2013).

The results of the first analysis which mapped out the trajectory types (low, stable, decreasing, increasing) for both internalising and externalising indicate that to a large extent individuals with low symptom trajectory in one domain have low symptom trajectories in the other. Although for all other trajectories symptom trajectories in internalising did not map onto the same symptom trajectory in externalising, however, there is a certain amount of overlap and a few patterns can be discerned. In the primary school sample almost all individuals with increasing internalising symptoms also had increasing externalising symptoms, whereas contrarily in the secondary school sample this was the case for a comparatively smaller proportion of individuals (17%). A high proportion of children with stable internalising symptoms had either low or an increasing externalising symptom trajectory whereas, in the older sample, most adolescents with a stable internalising symptom trajectory had low or decreasing externalising symptoms. On the other hand in the older sample a third having stable externalising symptoms had increasing internalising symptoms.

In terms of decreasing symptom trajectories there was no clear patterns in mapping in the younger sample but in the older sample a third of individuals with decreasing externalising symptoms had increasing internalising symptom whereas a third of individuals with decreasing internalising symptoms also had decreasing externalising symptoms.

The observed overlaps between types of symptom development trajectories in the present samples of children and early adolescents result in the following summary of findings:

1) A high proportion of children and adolescents who report low symptoms over time in either externalising or internalising domains, also report low symptoms in the other domain.
2) There are observable distinct patterns in how increasing, decreasing or stable symptom trajectories in one domain are associated with symptom development in the other domain and these relationships differ in the different age-groups.

Associations between baseline symptoms and symptom development

The second analysis identified latent intercept and slopes with the main aim of assessing the amount of association between symptom development of internalising and externalising. The analysis demonstrated an expected amount (Achenbach & Rescorla, 2001) of moderate co-morbidity between the baseline scores for internalising and externalising (primary r=.41; secondary r=.34). The correlations between slopes that symptom development, is also associated moderately (primary r= .37; secondary r=.36). Although these correlations are not very high, they are large enough to suggest that co-morbidity in symptoms is not only seen cross-sectionally, but that there is a similar amount of co-morbidity in the development of symptoms.

The relationships between cross-domain intercept and slopes were a bit more inconsistent across the age-groups and varied when slopes were predicted while controlling for same and other domain intercepts in the regression analysis. In the younger sample the

baseline internalising scores did not significantly predict slope of externalising symptoms, whereas baseline externalising predicted both internalising and externalising slopes.

In the older sample the intercept in each domain predicted its slope with baseline internalising scores significantly associated with development of internalising symptoms and baseline externalising scores significantly associated with development of externalising symptoms. Baseline scores in each domain were not significantly associated with development of symptoms in the other domain suggesting that to the larger extent internalising and externalising predict development of symptoms in the same domain (homotypic continuity) compared to symptom development in the other domain. However, when controlling for same domain intercept to examine impact on slope, other domain intercept significantly predicted the slope in both internalising and externalising. The magnitude of the co-efficients of same and other domain intercepts predicting slope for both internalising and externalising were similar, albeit in opposite directions (B ranging from -.04 to +.04). The negative correlation between the intercept and slope of the same domain indicates that individuals with high initial levels of symptoms were more likely to show lower growth or a decrease in symptoms, which is similar to findings reported by other studies in this age group that have used similar methodology (Measelle et al., 2006). The positive co-efficients of cross-domain intercept and slope predictions seem to indicate that baseline scores in each domain predict increases in symptoms in the other domain, with higher scores predicting greater symptom development. This is an interesting finding that can help clarify the longitudinal associations between internalising and externalising symptoms over time in these age groups. For instance, (Moilanen et al., 2010) using a cascade modelling approach, found that early externalising predicts increases in internalising symptoms, but not vice-versa and the results of the present study are in line with expectations from studies of sequential co-morbidity that suggest that in childhood externalising problems might be pre-cursors to internalising symptoms (Cerdá et al., 2008). However, as Cerdá et al. (2008) noted, the

evidence so far for internalising symptoms predicting externalising symptom development is sparse, which the results of the current study support but only in early adolescence. The findings of the current study suggest that this relationship is complex and varies by age/developmental stage. Similar analyses of co-development of internalising and externalising symptoms in childhood are not wide-spread and further investigation using such an approach might help uncover these longitudinal cross-domain associations at different stages of development.

Strengths and Limitations

The strengths of the current study include the possibility of examining co-development of symptoms in two cohorts (age 8-11 years and 11-14 years) which permits comparisons of different types of relationships at these two different developmental stages. The main aim of the present investigation was to map out overlap/co-occurrence of person-centered trajectories in internalising and externalising symptoms over a short time frame, which was achieved. However, the key limitation is the inability to examine the associated correlates of the distinct patterns in co-development of externalising and internalising trajectories. The lack of power in some trajectory groups coupled with the moderately sized sample limits the possibility of splitting of the sample into groups based on for example, gender or deprivation, to conduct comparisons. Moreover, this is compounded by the lack of an established methodology by which comparisons of predictors of group membership across so many possible combinations can be made. A possible approach for future analysis might consider a priori trajectory selection based on and assessment of change/no-change in symptoms using such an index such as reliable change in symptoms to then assess co-occurring trajectories of symptoms.

Implications and future directions

Both these analyses in this study indicate that, although internalising and externalising dimensions of child psychopathology are to the greater extent studied as separate distinct

entities; there is a moderate level of association or overlap both in the cross-sectional levels and longitudinal symptom development in children and adolescents. The findings regarding moderate longitudinal and cross-sectional overlap between the domains and their development lend support to the hypothesis that common aetiological factors underlie both types of disorders. This is also supported by emerging findings from behavioural genetic ACE models where the shared genetic risk associated with internalising and externalising disorders seem to be more similar than distinct (Lahey et al., 2011).

Revisiting the proposed explanations that exist for co-occurrences of symptoms and their development that were outlined in the discussion: shared risk, co-occurring risk, common syndrome and pathogenetic co-morbidity, the results of this and many studies also lead to an alternate proposal. It has been suggested that there might be a single 'core' or common factor that explains symptoms and symptom development in psychopathology (Measelle et al., 2006). This hypothesis, can be considered an extension of the hypothesis by Caron & Rutter that co-morbid patterns might be representing a meaningful syndrome or dimension. The possibility of a single broad dimension of psychopathology has the potential to encompass all previous hypotheses as it would explain shared risk, co-occurring risk and possibly to some extent sequential co-morbidity. Weiss, Süsser, and Catron (1998) suggest that childhood disorders could be conceptualised as having three levels of generality or specificity as follows 1) *common features,* which mainly serve to distinguish internalising and externalising from normality, 2) *broadband-specific features,* distinguishing internalising and externalising from each other and 3) *narrowband-specific features*, that serve to distinguish between disorders within the internalising or externalising domains.

However, studies that have tried to fit a single factor to child psychopathology data have usually found that one factor is poorer fit than the traditional two-factor internalising externalising (Measelle et al., 2006) or multiple factors with specific disorder categories. These studies are limited in the fact that when they consider the one general factor they do not

simultaneously account for the specific internalising and externalising factors (Measelle et al., 2006). In the Weiss et al. (1998) theoretical model the general factor does not replace the 2-factor internalising and externalising but rather is present in a hierarchical structure above them. To this effect, similar alternative ways of conceptualising the underlying structure of psychopathology have recently been analysed and proposed by researchers exploring the structure of adult psychopathology. These models include a hierarchical bi-factor which demonstrates the possibility of a general psychopathology bi-factor while retaining the specific internalising, externalising and thought disorder dimensions in adult psychopathology (Caspi et al., 2013; Lahey et al., 2012). These models, not only identified a general factor of psychopathology, but demonstrated that the internalising and externalising dimensions remaining after the general bi-factor is accounted for might be associated with more specific risk factors. Future investigations of child psychopathology could consider these alternative approaches to exploring the structure of psychopathology and the existence of concurrent and sequential co-morbidity.

Study 6. A general psychopathology factor in early adolescence?

Structure, correlates and predictive utility

Background

In child psychopathology, two main dimensions have been widely used for decades to characterize the structure of mental health disorders under the age of 18: most commonly referred to as Internalising and Externalising (Achenbach, 1966; Achenbach & Edelbrock, 1978). They have been regarded as distinct and key dimensions of child mental health difficulties and a bulk of research findings studying aetiology, correlates, development and classifications of child and adolescent psychopathology are considered in relation to these two distinct dimensions. Evidence supports the existence of both age and gender related patterns in relation to these dimensions; thus externalising disorders are more likely to be early onset and afflict males while adolescence and females is characterised by greater prevalence of internalising symptoms (Martel, 2013).

However, the dimensions are associated moderately, with studies indicating correlations between .4-.6 (Achenbach & Rescorla, 2001; Deighton, Tymms, et al., 2013), indicating a certain amount of co-morbidity is present between them. This association is seen in cross-sectional examinations of child psychopathology, and it can be observed in longitudinal studies of development of symptoms (Measelle et al., 2006), findings which have been summarised in the previous studies, the previous study providing another example of the moderate associations of internalising and externalising symptoms (Study 5). Other studies using different approaches, for instance cascade modelling, have also observed a certain dependence or cyclical relationship between these two dimensions over time (Masten et al., 2005). This simultaneous co-occurrence or sequential co-morbidity of disorders in both childhood and adult psychopathology has led to many questioning the categorical classifications of disorders, as is common practice in current nosologies of mental disorders such as the DSM and ICD (Caron & Rutter, 1991; Cerdá et al., 2008).

Of relevance or interest, in terms of aetiology, development and treatment, is the risk factors associated with these dimensions and their environmental, genetic and neurological

correlates. However, decades of research into risk factors has hardly provided us with conclusive evidence on risk factors associated uniquely with either internalising or externalising; gender being the clear exception where males are more likely to externalize and females to internalize (Martel, 2013). Instead, what is observed in the literature is identification of risk factors that are more similar than different for the two domains. For instance, deprivation is linked to both internalising and externalising disorders (Green et al., 2005), although some evidence suggests stronger links with externalising (Murray & Farrington, 2010); childhood maltreatment is a risk factor that predicts many different psychiatric disorders including internalising, externalising and psychosis in adulthood (Green et al., 2010; Varese et al., 2012); parental psychopathology also predicts a variety of childhood and adulthood diagnoses (Kessler et al., 2010). This pattern of non-specificity is also observed in recent attempts to identify the neurological or genetic correlates of these two dimensions. For instance, a twin study in children and adolescents tried to partition genetic influences associated with internalising and externalising and found that a general genetic risk bi-factor best explained prevailing psychopathology in twins and suggested a genetic liability that functions across all disorders (Lahey et al., 2011). Similarly, attempts to observe specific gene correlates of specific disorders has so far been unsuccessful, for example, the *Cross-Disorder Group of the Psychiatric Genomics Consortium* tried to identify specific gene loci associated with disorders and conclude that specific SNP's are associated with a range of childhood and adulthood disorders (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013). Caspi et al. (2013) examined neurological correlates and brain integrity and concluded that lower brain integrity was associated with the general propensity towards psychopathology rather than specific disorder types. Moreover, this pattern of non-specificity is also observed in how these dimensions are associated with other domains of childhood functioning, for instance educational attainment (Studies 3&4).

In the light of what we know today about the risk factors and correlates (environmental, genetic, neurological) of the two key broad dimensions of child psychopathology attention is now turned to the dimensions themselves, which is discussed in the next section.

Structure of child psychopathology

As briefly mentioned above, the two main dimensions underpinning child psychopathology for many decades have been the internalising and the externalising dimensions. This approach for categorising childhood disorders was first popularised by Thomas Achenbach through the 1960's and 70's (Achenbach, 1966; Achenbach & Edelbrock, 1978) and since have been the mainstay of childhood psychopathology, especially with regards to research where this simple nosology is used more commonly used than in practice where more detailed nosologies are used (e.g., ICD-10). The internalising spectrum generally encompasses depressive symptoms, anxiety, phobias and externalising includes conduct disorder, operational defiant disorder, alcohol and substance misuse. Although, the existence of internalising and externalising dimensions has been established and is widely accepted, attempts at explaining co-morbidity of externalising and internalising have resulted in re-explorations of these dimensions in child psychopathology. For instance, (Weiss et al., 1998) suggest that childhood disorders could be conceptualised as having three levels of generality or specificity as follows 1) *common features,* which mainly serve to distinguish internalising and externalising from normality, 2) *broadband-specific features,* distinguishing internalising and externalising from each other and 3) *narrowband-specific features,* that serve to distinguish between disorders within the internalising or externalising domains. They suggested that this approach to conceptualising child psychopathology might permit understanding co-morbidity and that the common features might be related to severity of disorder and the specific features would then be related to differentiation of psychopathology.

In adult psychopathology, on the other hand, such clear, long standing, distinct dimensions such as internalising and externalising have not existed nor endured over so many decades as they have in the study of child psychopathology. Analyses and debates into the structure of psychopathology have been occurring to greater or lesser extent over many decades. The latent structure underpinning psychiatric diagnoses in adults has been a subject of recent renewed interest (Caspi et al., 2013; Kotov et al., 2011; Lahey et al., 2012; Wright et al., 2013). In particular, the existence of similar correlations and co-morbidity between disorders and symptoms in adult populations has led to recent re-examinations of the dimensionality of psychopathology in two key studies, one of 18-64 year olds (Lahey et al., 2012) and the other of participants in a longitudinal study repeatedly assessed at 18, 21, 26, 32, and 38 (Caspi et al., 2013). These two studies assessed the existence of a bi-factor or hierarchical factor by analysing data using bi-factor approaches. Both studies concluded that a hierarchical model fit the data best, one in which (1) all the various dimensions loaded either onto distinct externalising, fears and distress (Lahey et al., 2012) or internalising and externalising (Caspi et al., 2013) dimensions at one level, yet (2) onto a single, dimension at another level. Caspi and associates (2013) refer to the higher-order dimension as general psychopathology or '*p*', drawing structural and conceptual parallels to the general intelligence factor '*g*'. The results of both studies indicate that *p* better represents longitudinal risk of psychiatric disorders and suggest that studying this factor will allow for a better understanding of aetiology, correlates and prognosis of psychiatric disorder. The risk factors examined in these studies indicate that external correlates are linked with the internalising, externalising and thought disorder dimensions primarily because they are associated with *p*. The only exception was gender which was not significantly associated with *p* suggesting the general propensity to experience psychopathology might be equal across genders (Caspi et al., 2013).

The current study

Based on these investigations into the structure of adult psychopathology, it seems conceivable that a general psychopathology factor might characterize child mental health as well. To address this issue, the current study examined the structure of psychopathology in a large community-based sample substantially younger than those included in the aforementioned adult work: 11-14 year olds. The goal was to ascertain whether a general psychopathology dimension could be identified in early adolescence. Moreover, the study sought to evaluate the external validity and relevance of a hierarchical model that includes *p* by exploring associations with socio-demographic and educational correlates known to be associated with child psychopathology (Green et al., 2005). Additionally, the ability of the derived dimensions to predict future psychopathology one year later was assessed, thereby determining the predictive utility of a general, bi-factor dimension of psychopathology.

Method

Participants

Data were collected from 23,477 participants (50.4% female) across 210 state-maintained secondary schools in England as part of a national study of mental health and provision in schools, the Me and My School study. Details of the wider study are published elsewhere (Wolpert et al., 2011). Mean age was 12.05 (SD=.56), with 99.9% of the sample ranging from 11 to 13.5 years of age. The sample was predominantly White (76.2%), followed by Asian (8.7%), Black (5.9%), Mixed (3.8), and other (1.4%) ethnicities; 2.9% were unclassified. One of five (19.7%) were eligible for free school meals which serves as a proxy for economic deprivation.(Hobbs & Vignoles, 2010) The socio-demographic attributes of the sample reveal it to be broadly representative of the general population early adolescents (DFE, 2010b).

Follow-up data were available for:

(1) Psychopathology: a sub-sample of 10,270 participants (Mean age= 13.07 [.56], 51.6% female, 18.5% eligible for FSM, 76.5% White) from 124 schools completed the mental health measures at a follow-up assessment a year on.

(2) Academic attainment: a sub-sample of 7,569 participants who were in Year 8 at time 1 (54.4% female, 16.6% eligible for FSM, 77.6% White), had standardised national test scores a year and a half later.

Procedure

Computer-based surveys were completed by pupils within the normal school day with support provided by teachers. More specifically, teachers read to the class standardised information, including about the study, confidentiality of responses, and the right not to participate or drop out at any time. Data on socio-demographic characteristics such as gender,

SES, ethnicity and age were obtained from the national pupil database, a centrally collated database that holds all education related data on all pupils in England.

Measures

Psychopathology

Participants completed two questionnaires reporting on mental health symptoms, Me and My School (M&MS) and the Strengths and Difficulties Questionnaire (SDQ). Items from both measures were used because the measures have quite distinct items and using multiple measures was judged to increase validity, especially as the factor analyses to be reported were not meant to represent structures of existing 'measures' but instead of the 'constructs/dimensions' of psychopathology that they represent. The only similar item across the two measures was 'I worry a lot'; it was retained in both as they correlated only .55 across the two instruments. This probably resulted from the different response options of the two measures: the frequency-based response options of the M&MS and the endorsement-based options of the SDQ (see below).

Me and My School questionnaire (M&MS). Symptoms were measured using the 10-item emotional difficulties and the 6-item behavioural difficulties scales of the Me and My School questionnaire (Deighton, Tymms, et al., 2013). This community based screening measure of mental health difficulties has relatively simple, easy-to understand items and has been validated for use by children from 8 years of age. Participants respond to each item by endorsing one of three response options: never, sometimes, always. Preliminary analysis revealed that one item, 'I am shy', that belonged to the emotional difficulties scale loaded poorly onto the internalising scale (factor loading = .03) which led to this item being excluded from the final analysis. The internal reliability, Cronbach's alpha, was .78 for the emotional difficulties and .80 for the behavioural difficulties scales in the current sample.

Strengths and Difficulties questionnaire (SDQ). The SDQ is a widely used self-report measure of child mental health.(Goodman et al., 1998) The five-item emotional symptoms and conduct problems scales were used in the present study and participants respond to each item by endorsing one of three response options: not true, somewhat true and certainly true. Cronbach's alpha was .72 for the emotional symptoms and .66 for the conduct problems scales in the current sample.

Correlates

The external variables included in this study include gender, socio-economic status, educational attainment and special educational needs (SEN). Socio-economic status was measured using free school meal eligibility (FSM) which is a binary indicator often used as a school based proxy for deprivation and the Income Deprivation Affecting Children Index (IDACI) which is a variable representing the deprivation ranking of the neighbourhood in which a child lives. SEN was based on extent of special educational provision for each student and had the values no SEN, school action, school action plus and statemented. Educational attainment was assessed using standardised national assessment scores averaged across English, mathematics and science. For detailed information on the correlates please see measures in Study 3 (pages 122-123).

Future functioning

Psychopathology. Self-reported mental health symptoms in the follow-up wave a year on were used to group participants based on case-ness or not based on both the M&MS and SDQ scores. SDQ total difficulties was based on the abnormal threshold score of 20. M&MS emotional symptoms used the clinical threshold of 12 and behavioural symptoms score of 7 with M&MS overall case-ness indicated by above threshold scores on either scale. For all the variables individuals above threshold were coded '1' and below threshold '0' for the analysis. From the sub-sample of 10,270 participants who completed the M&MS and SDQ a year on, based on the M&MS clinical threshold, 11.7% were classified as exhibiting case-ness at this

time point, with 3.9% above threshold on the emotional difficulties and 9.4% on the behavioural difficulties scale. For the SDQ total difficulties, 7.3% had above threshold scores at follow-up.

Poor academic attainment. National standardised test scores taken at the end of Year 9 in England, referred to as Key stage 3, were used as measures of academic functioning. Scores can range from 0-8 and according to government set standards, pupils are expected to achieve at least level 5.[16] As outlined in the participants section, Key stage 3 scores were available for the 7,569 participants who were in Year 8 at time 1. The variable was coded '1' for 1,832 participants (24.2%) who had achieved below level 5, indicating poor academic functioning.

Analytic strategy

Analyses were conducted in three stages (1) to examine the structure of psychopathology and the possibility of a single, hierarchical dimension of general psychopathology factor in young people, (2) to evaluate the socio-demographic and educational correlates of the (a) general, (b) internalising and (c) externalising factors in the alternate factor solutions from stage 1, and 3) to determine the predictive value of the different dimensions by assessing the extent to which *p* and the internalising and externalising dimensions predicted future psychopathology assessed a year on.

Stage 1: The structure of psychopathology

Several constructs in psychology can be considered hierarchical, where different levels of the constructs operate at different levels of generality (Emmons, 1995), resulting in higher order general factors and nested specific factors. For instance, research into the structure of well-being suggests a hierarchical structure whereby well-being is the general bi-factor and eudaimonic, hedonic and social well-being are the specific factors (Gallagher, Lopez, & Preacher, 2009). In this study we utilised the standard approach to establishing hierarchical structured dimensions or constructs (Brunner, Nagy, & Wilhelm, 2012), which is similar to the approach used by Caspi et al.(2013) in their study in adults. Confirmatory Factor Analyses (CFAs) were conducted to test three different models, (1) a correlated factors model, (2) a bi-factor or hierarchical model, and (3) a 1-factor model. The first model concerns the widely used 2-factor, internalising and externalising factor solution. The second model introduces the hierarchical dimension of a general psychopathology bi-factor or 'p', onto which every manifest variable loads, in addition to loading onto the internalising or externalising dimension that they represent. The third model is a 1-factor model in which all items load onto a single factor; it evaluates the hypothesis that the single, 1-factor model can account for variation without the need for specific lower-order internalising and externalising factors. Results from Caspi et al.(2013) suggest this third model is not a good fit to the data; given the

established nature of internalising and externalising factors in child psychopathology, we expect this model not to fit the data well in our data.

CFA's were estimated in Mplus 7 (Muthén & Muthén, 2012) and the weighted least square means and variances (WLSMV) estimator was used as it is the most suited for categorical manifest variables (Finney & DiStefano, 2006). Model fit was assessed using the comparative fit index (CFI), Tucker Lewis index (TLI) and root mean square error of approximation (RMSEA). CFI and TLI closer to 1 indicate good fit to the data and RMSEA closer to 0 indicates good fit. The values were judged by the widely used conservative criteria of CFI and TLI greater than .95 and RMSEA less than .06 to indicate very good model fit (Hu & Bentler, 1999). Due to the non-nested nature of the models, models could not be compared directly using a difference in fit statistic (e.g., Akaike information criterion) being based on maximum likelihood estimation (Bozdogan, 1987), they cannot be estimated when WLSMV estimation is used. Hence, model fit criteria and factor loadings were used to assess the quality of models.

Factor scores for every latent dimension from each model were outputted and used in the following stages of analysis.

Stage 2: Associations with external correlates

The associations between derived factor scores for the dimensions of interest and the external correlates were examined in this stage of analysis. Factors scores from different models computed in Stage 1 were correlated with variables with established associations with psychopathology in childhood in order to assess the external validity of the dimensions while examining the relative associations of *p* and the specific internalising and externalising factors with the external correlates. Pearson's correlations were computed where correlates were continuous variables and Spearman's correlations were computed for the categorical variables.

Stage 3: Predicting future functioning

In this stage the predictive capacity of scores from resulting dimensions to predict psychopathology and academic attainment were examined. Psychopathology scores from a year later were dichotomised to reflect case-ness based on clinical cut-offs of the measures into above and below cut-off (above=1, below=0). Similarly, poor academic performance was indicated by a binary variable (poor attainment=1). Logistic regressions were conducted with the factor scores from the alternative models predicting case-ness. Regression co-efficients for the factor scores are reported alongside odds-ratios indicating the odds with which the dimension scores predict functioning a year later.

Results

Stage 1: The structure of psychopathology

Results for the following three CFA models computed in this stage are presented below, 1) 2-factor model, 2) bi-factor model and 3) 1-factor model and factor loading and model fit statistics for the three models are presented in Table 6.1 and 6.2.

Model 1: 2-factor model with internalising and externalising dimensions

The first CFA tested the commonly employed 2-factor model with separate internalising and externalising dimensions. The model allows for the internalising and the externalising factors to be correlated. The first columns in Table 6.1 presents the model fit statistics and Table 6.2 contains the standardised factor loadings for this model. As might be expected, factor loadings of the items on the two factors were all positive, above .5 and significant ($p<.001$), indicating that the 2-factor model explains the data reasonably well. Moreover, model fit statistics revealed that the model-fit was acceptable; RMSEA was .06 and CFI (.93) and TLI (.93) were just under the accepted threshold for very good model fit (.95).

Table 6.1. Model fit statistics for the three models

Model fit statistics	**Model 1: 2-factor model**	**Model 2: bi-factor model**	**Model 3: 1-factor model**
TLI	0.93	0.94	0.68
CFI	0.93	0.95	0.7
RMSEA (90% CI)	.060 (.059-.060)	.051 (.051-.052)	.124 (.123-.125)
$\chi 2$ (df)	23097.19 (274)	15723.15 (250)	99715.84 (275)

Table 6.2. Standardised factor loadings from the three models

Items	**Model 1: 2-factor model**		**Model 2: bi-factor model**			**Model 3: 1-factor model**
	Internalising	*Externalising*	*Internalising*	*Externalising*	*p*	*p*
I feel lonely	.62		.44		.44	.52
I am unhappy	.62		.35		.51	.54
Nobody likes me	.56		.33		.46	.48
I cry a lot	.63		.47		.44	.53
I worry at school	.72		.56		.47	.62
I worry a lot	.71		.73		.33	.59
I have problems sleeping	.58		.07		.67	.51
I wake up in the night	.54		.00*(ns)*		.66	.48
I feel scared	.68		.53		.45	.57
I get a lot of headaches, stomach aches, sickness	.52		.15		.53	.46
I worry a lot	.80		.73		.43	.69
Often unhappy, downhearted, tearful	.78		.44		.65	.69
Nervous in new situations. Easily lose confidence	.51		.35		.38	.43
I have many fears, I am easily scared	.64		.51		.41	.53
I get very angry		.84		.64	.53	.78
I lose my temper		.87		.72	.47	.80
I do things to hurt people		.73		.63	.37	.63
I am calm		.59		.46	.35	.51
I hit out when I am angry		.82		.76	.37	.73
I break things on purpose		.64		.57	.31	.53
I get very angry and often lose my temper		.85		.72	.45	.78
I usually do as I'm told		.56		.61	.16	.44
I fight a lot. I can make other people do what I want		.67		.65	.26	.55
I am often accused of lying and cheating		.59		.39	.44	.52
I take things that are not mine from home, school or elsewhere		.50		.38	.31	.42

Note: *all factor loadings except the one marked (ns) were statistically significant at at-least the .05 level.*

Model 2: Bi-factor model with a general psychopathology bi-factor

The second model was a bi-factor model with internalising and externalising as lower-order factors and p as a higher order bi-factor (see Figure 6.1); this model assumes that the derived factors are not correlated (Yung, Thissen, & McLeod, 1999). This model fit the data well (CFI= .95, TLI=.94, RMSEA=.05; see Table 1 for factor loadings). Loadings on the general factor were all moderate and significant (p<.001), with an average factor loading of .43. However, in this model loadings on internalising were not all high and significant, with both items relating to sleep ('I wake up in the night'; 'I have problems sleeping') having zero order loadings (.07, .0002) on the internalising-specific factor. This suggests that these two items more directly predict general psychopathology than internalising. Factor loadings on the externalising factor were all moderate-high and significant (p<.001), with an average factor loading of .59. Based on model fit statistics, this bi-factor model seemed to be a slightly better fit to the data than the 2-factor model.

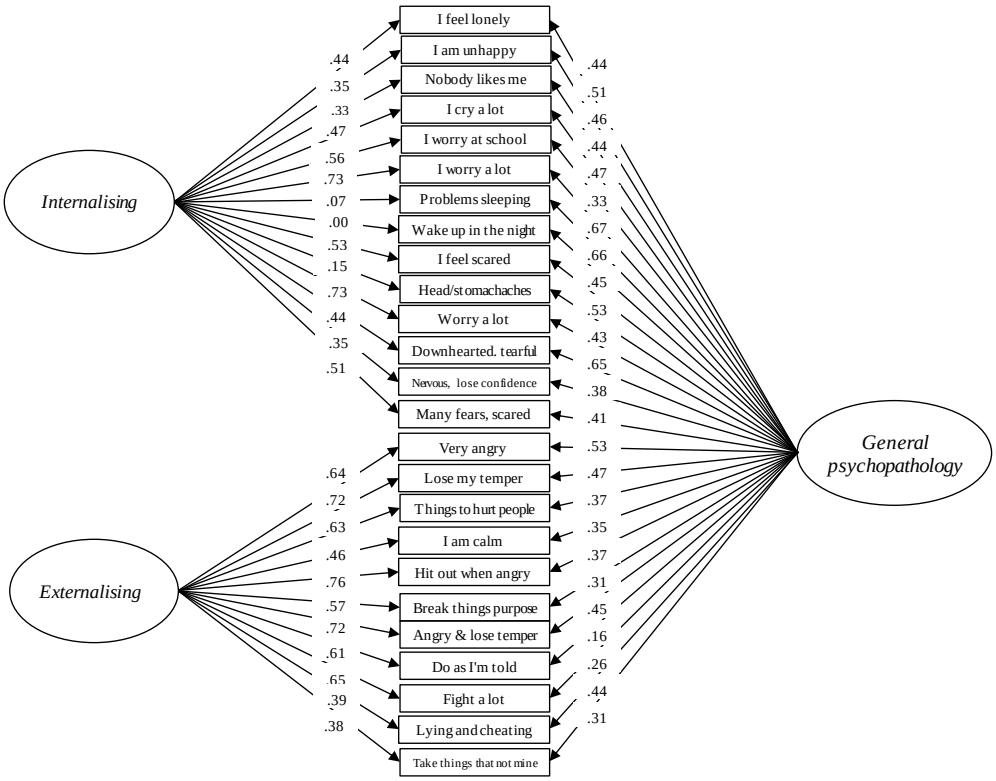


Figure 6.1. Bi-factor model with the item-loadings onto the internalising and externalising dimensions and the general psychopathology bi-factor

Model 3: 1- factor model

The 1-factor model assigned each item only to a general overall factor. Model fit statistics indicate that this model did not fit the data well (CFI= .7, TLI=.68, RMSEA=.12, see Table 6.1 for factor loadings). Factor loadings on the 1-factor were moderately high and significant (p<.001), with an average loading of 0.57.

To summarize, the first set of analyses indicate that the hierarchical bi-factor solution explains the data best, although the traditional internalising-externalising 2-factor model proved almost as good. The 1-factor model did not explain the data well, clearly suggesting that a general psychopathology factor on its own is not sufficient. Therefore, subsequent analysis involving established correlates/predictors of psychopathology only focussed on factors from the first two models. For this purpose, factor scores were outputted from both models (1&2).

Correlations between these factor scores were estimated in both models. In the 2-factor model the correlation between internalising and externalising was moderate (.45, p<.001), whereas in the bi-factor model internalising and externalising correlated negatively and more weakly (-.16, p<.001). The general psychopathology bi-factor *p* correlated .30 (p<.001) with internalising and .22 (p<.001) with externalising scores in the bi-factor model.

Stage 2: Associations with external correlates

Correlations (either Spearman's or Pearson's as appropriate) with the factor scores from Models 1&2 above are presented in Table 6.3.We can see from Table 6.3 that gender did not correlate significantly with p in the bi-factor model. Girls scored significantly higher, though, on internalising and significantly lower on externalising than boys in both the 2- and bi-factor models. Correlations involving both indicators of social class, FSM and IDACI,

revealed that lower social class was related to higher levels of p and greater externalising problems in both models. In the case of internalising, however, the significant association between internalising and the two SES indicators proved positive in the 2-factor model but negative in the bi-factor model; in absolute terms, however, these significant correlation coefficients did not differ by very much, as each hovered around zero, proving significant due to the large sample size. Finally, children receiving SEN classifications scored higher on problems, however parameterised, in both models, with the reverse being true of children who scored high on attainment. In the latter case, however, the association with internalising problems in the bi-factor model was not significant.

Table 6.3. Correlations between factor scores and predictors

Predictor	**2-factor model (Model 1)**		**Bi-factor model (Model 2)**		
	Internalising	Externalising	Internalising	Externalising	*p*
Gender$^{\varrho}$ (Female)	.13**	-.21**	.23**	-.27**	-.007
FSM$^{\varrho}$ (Yes)	.04**	.14**	-.02**	.14**	.08**
IDACI	.02*	.14**	-.05**	.14**	.08**
SEN$^{\varrho}$ (Yes)	.10**	.14**	.03**	.11**	.13**
Attainment	-.1**	-.2**	-.001	-.17**	-.14**

Note: $^{\varrho}$ indicates non-parametric correlations (Spearman's rho). *p< .01, **p<.001. FSM= Free school meal eligibility, IDACI=Income deprivation affecting children index, SEN=Special educational needs

Stage 3: Predicting future psychopathology

To evaluate the predictive validity of the factors scores derived from the 2-factor and bi-factor models, logistic regressions were conducted to predict future psychopathology case-ness and academic attainment using factor scores from both models, while controlling for socio-demographic correlates including gender, SES and ethnicity (results in Table 6.4). From the bi-factor model the regression co-efficients indicated that all three predictors significantly and positively predicted future psychopathology measured with both the M&MS and SDQ. Odds-ratios (ORs) not only suggest that the general psychopathology dimension predicted future psychopathology over and above that of the other two dimensions, (M&MS,

OR=10.08; SDQ, OR=17.05), but did so to a greater extent than them (internalising, ORs=1.27 & 2.63, externalising, ORs=4.04 & 2.59). Factor-specific scores from the 2-factor internalising-externalising model predicted future psychopathology to a similar extent as the specific scales from the bi-factor model (internalising, ORs=1.83 & 4.23, externalising, ORs=3.97 & 2.72). Even in the case of specific emotional or behavioural symptoms the general psychopathology factor was the best predictor of future symptoms (predicting emotional, OR=20.54, predicting behavioural, OR=7.25).

Regressions predicting future academic functioning indicate that in the bi-factor model the general psychopathology factor (OR=2.34), externalising (OR=1.91) and internalising (OR=1.32) significantly predicted future attainment. Both the internalising (OR=1.33) and externalising (OR=1.76) dimensions from the 2-factor model significantly predicted future academic functioning to a similar extent as the corresponding dimension from the bi-factor model.

Based on an effect size interpretation, where OR $\geq$ 6.71 is considered a large effect (Chen, Cohen, & Chen, 2010), *p* predicting future psychopathology were large effects, with most other domains having small or medium predictive capacity.

Table 6.4. Logistic regressions predicting future functioning

Predictor	**SDQ total difficulties**		**M&MS overall**		**M&MS emotional**		**M&MS behavioural**		**Academic attainment**	
	B	OR	B	OR	B	OR	B	OR	B	OR
2-factor model										
Internalising	1.44***	4.23	.60***	1.83	2.27***	9.65	.04	1.04	.29***	1.33
Externalising	1.00***	2.72	1.38***	3.97	0.27**	1.28	1.72***	5.58	.57***	1.76
Bi-factor model										
Internalising	.97***	2.63	0.24*	1.27	1.61***	5.00	-0.19	0.83	.28**	1.32
Externalising	.95***	2.59	1.39***	4.04	0.08	1.08	1.82***	6.16	.65***	1.91

p	2.84***	17.05	2.31***	10.08	3.02***	20.54	1.98***	7.25	.85***	2.34

*p< .05, ** p<.01, ***p<.001 Note. Sample size for psychopathology measures n=10270 and academic attainment n=7569

Discussion

In recent years there has been resurgence of interest in trying to understand and simplify the dimensionality of adult psychopathology (Kotov et al., 2011; Wright et al., 2013). Thus, Lahey and colleagues (2012) first tested and confirmed the existence of a higher dimension, general-psychopathology factor which represents what different disorders share in common. This structure was subsequently verified by Caspi et al. (2013) in adults. Both investigations indicate that this general psychopathology factor was independent of other dimensions and was associated with known external correlates of disorder that predict future psychopathology. As the youngest participants in both of these studies were 18 years of age, we sought to determine whether a similar psychopathology structure would emerge with an adolescent sample. We also examined the external correlates and predictive validity of dimensions of psychopathology discerned. Results of each set of analyses are discussed in turn, followed by a discussion of implications and future directions.

A general psychopathology dimension in young people

Results with our younger sample proved very much in line with those emerging from the two cited studies of general psychopathology in adults (Caspi et al., 2013; Lahey et al., 2012). The results of the current study also found that a hierarchical model reveals a higher dimension p factor over and above the classic 2-factor, internalising and externalising dimensions. Comparison of the bi-factor model and the traditional 2-factor model yielded some interesting insights into the structures and relationships between these dimensions (or disorder liabilities). With regard to associations between internalising and externalising dimensions, the traditional 2-factor model yielded a correlation of .45 which is in-line with the known moderate association between these dimensions. However, in the bi-factor model, internalising and externalising correlated negatively (r= -.16), although the strength of the negative correlation is lower, this suggests that previously identified associations between internalising and externalising are mainly explained by *p*. After removing variance associated

with general vulnerability to psychopathology, internalising and externalising emerge as distinct yet unrelated styles of expressing psychopathology with potentially unique relationships with other demographic and clinical characteristics.

Importantly, the factor loadings in the different models result in clearer understanding of the externalising and internalising dimensions at the symptom level once the general common vulnerability to psychopathology is removed. For instance, sleep disturbance is a poor indicator of internalising problems and might be better conceived of as a generic indicator of vulnerability to psychiatric disorder. All the other items, in most cases, represent the general factor to a certain extent and also the specific factor they are meant to represent. Considering the different age-group and geographical location of the current sample when compared to the two studies that focussed on adult psychopathology (Caspi et al., 2013; Lahey et al., 2012), as well as the use of symptom-level rather than diagnosis-level variables used in the current investigation, the findings reported here clearly buttress the claim emerging from the prior work that there exists a general psychopathology factor, one that is now discernible from at least the beginning of the second decade of life.

The single factor model, with only a general psychopathology factor, was a poor fit to the data, which would be expected based on both recent explorations in adults (Caspi et al., 2013) but also previous studies investigating the possibility of a single factor in adolescence (Garnefski & Diekstra, 1997). The results from this model stress the importance of retaining the existing internalising and externalising dimensions in studying the factor-structure of psychopathology.

External correlates

Examination of some socio-demographic and educational correlates of the dimensions in the two models further illuminated differences between them. In the 2-factor internalising-

externalising model all the examined correlates were associated significantly with the two dimensions, as would be expected from known associations of these variables in young people(Green et al., 2005). This situation changed, however, once the general psychopathology dimension, *p*, was taken into account.

Consider first the case of internalising problems, the internalising dimension had reduced associations with SEN and previously positive associations between internalising and indices of socioeconomic deprivation became smaller and negative once p was taken into account, while associations with educational attainment completely disappeared. The last finding might to some extent explain why research examining associations between internalising and education have been largely inconclusive (Masten et al., 2005). The results from the bi-factor model suggest that internalising problems carry educational and social problems only to the extent that they are linked with a general vulnerability to psychiatric disorder. The unique contribution of internalising problems, although significant in a large population based study, is almost negligible in terms of prediction to attainment or intervention to support educational need. This finding, if found to be robust in other future studies, may have significant policy implications in terms of school based screening, support and intervention for young people.

The externalising dimension, once *p* was taken into account, retained positive associations with the deprivation variables, indicating that greater deprivation might be linked uniquely with greater externalising even after general psychopathology is accounted for. Similarly, externalising post-p retained negative associations with educational attainment. These findings indicate that characteristics of externalising problems uniquely impact on learning and attainment, even after a general propensity to psychopathology is accounted for. However, it is important to bear in mind that the correlation co-efficients for some of these associations were small, significance being reached due to the large sample, replication and more detailed investigations into these associations are required before more definite

conclusions can be made. In any case, the methodology demonstrated and the preliminary results indicate the potential usefulness of such an approach to allow researchers to partition specific risk factors associated with internalising or externalising disorders.

In the current study gender was not associated with *p*, a finding similar to that reported in at least one of the prior studies of *p* in adults (Caspi et al., 2013). Additionally, after variance associated with general vulnerability was controlled, associations between gender and the internalising and externalising dimensions increased. This suggests that the gender specificity of the internalising and externalising spectra (frequently observed in epidemiological studies (Green et al., 2005) may be stronger than previously thought once overall vulnerability to mental disorder is separately identified. In the study of childhood psychopathology, gender has been a variable of much focus to the extent that it has become common practice to study single-sex samples (Moffitt, Caspi, Dickson, Silva, & Stanton, 2009) for both internalising and externalising disorders. In doing so, the study of co-morbid symptoms and other risk factors quite often take a back seat to gender in studies of developmental psychopathology. The general propensity for psychopathology might actually be equal across genders. Although it remains important to understand the differential inclinations to developing certain kinds of disorders, which is likely to be a result of a multiplicity of factors (e.g.,, hormones (Nolen-Hoeksema & Girgus, 1994), evolutionary pressures (Martel, 2013)), The finding that gender is not associated with a general liability to psychopathology suggests a need to reconsider traditional approaches to studying gender and psychopathology, including the role of gender in research on the aetiology and development of disorder. The findings reported here suggest that 'gender' should be a less important part of efforts to understand a general liability to psychopathology, yet an important focus in models of specific disorders and their associated correlates.

Predicting future psychopathology

In terms of predictive value, the general psychopathology bi-factor significantly predicted odds of future psychopathology over and above odds predicted by the internalising and externalising dimensions which, even if statistically significant, had much lower predictive value. For instance future case-ness as measured by the SDQ was predicted with OR of ~18 which internalising and externalising predicted future case-ness with ORs of only 2.5 and 2.8. It is also relevant that even specific future internalising or externalising were predicted better by *p* than by the specific internalising or externalising factors; for instance, future internalising was predicted with 1.57 OR by prior internalising compared to OR ~20 by *p*. These results are consistent with similar findings in the prior studies of older individuals and suggest that utilising the general psychopathology dimension can, as Lahey et al. (2012, p. 77) observed, *"substantially improve prognostic predictions of future psychopathology and functioning"*, now at least beginning as early as the second decade of life.

Strengths and limitations

It has been suggested that the general psychopathology dimension emerging in this and prior work might be a statistical artefact (Caspi et al., 2013) However, evidence of external validity and predictive utility from this and the two existing adult studies clearly suggest otherwise and support wider implications and applications of this bi-factor dimension. Lahey and colleagues (2012) note the possibility that the correlations between different disorder symptoms might in part be due to biases in reporting based on implicit theories of psychopathology. They suggest that some individuals experiencing one symptom might also report other symptoms based on expectations of symptoms rather than their actual occurrence. In child psychopathology research the practice of collecting proxy reported symptoms might to an extent help verify this hypothesis, although it is likely that proxy reporters might also demonstrate these biases. In any case replicating these analyses with proxy reported child psychopathology data would be interesting and might help advance understanding of the high levels of disagreement between reporters of child psychopathology (De Los Reyes, 2013).

The current study is limited in the conclusions that can be drawn because the items included in the measurement instruments do not represent the full range of psychopathology symptoms experienced in childhood and adolescence. However as recognised by both the adult studies (Caspi et al., 2013; Lahey et al., 2012), and even more pertinent to child studies is the lack of population level datasets that have measures across all possible diagnostic categories of psychopathology. Replication with a variety of different samples and a more comprehensive set of measures may be one way of validating and increasing our understanding of the general psychopathology dimension or liability towards disorder.

Implications and future directions

Caspi et al. (2013) also point out that although thought disorders are a distinct third category in adult psychopathology, they are absent in the study of child psychopathology. In the current study the two variables related to difficulties in sleeping loaded exclusive onto *p* which suggests that these variables might be indicators of a precursor to the thought disorder dimension in adulthood. However, the item on somatic symptoms (headaches, stomach aches, sickness) also loaded predominantly on *p* rather than internalising; this suggests the alternate possibility that they, along with the sleep items, might represent something more organic or represent a psychosomatic aspect of general psychopathology, which is reflected in their separate categorisation in diagnostic schema. It is a limitation of the current study that the primary measures used were developed with a 2-factor model of psychopathology in mind. In various preliminary models of instruments, cross-loading items between the two spectra of externalising and internalising may well have been selectively eliminated to improve fit. Future studies may need to create instruments deliberately designed to capture the hierarchical bi-factor model.

Caron and Rutter (1991), in support of studying co-morbidity in psychopathology, argued that studying co-morbidity was important as by ignoring it misleading conclusions could be drawn where a study of a disorder might produce results that are largely a

consequence of another co-morbid disorder. Based on the current study similar caution can be drawn but instead regarding general psychopathology. Studying this general risk or propensity to experience psychopathology is necessary and important as consequences and risk factors now attributed to certain disorders might actually be a cause/consequence of the general psychopathology factor.

Finding a general psychopathology factor at a younger age provides support for the hypothesis that diagnostic specificity increases with age. Childhood disorders tend to be less clear-cut into diagnostic criteria when compared to adult disorders (e.g anxiety and depression are less distinct in young people (Moffitt et al., 2007)), which is likely to be one of the reasons the broader classifications of internalising and externalising dimensions have been long established and employed in child psychopathology research (Caspi et al., 2013). The possibility of liability to general psychopathology even in childhood suggests that disorder specificity increasing with age might be a result of gradually increasing tendencies to express psychopathology in certain ways. It will be of interest to determine whether a general *p* factor would emerge in research on children during the elementary or even preschool years.

The existence of a general psychopathology deficit also has implications for the study of co-morbidities of disorder. Aseltine, Gore, and Colten (1998), recognised that studying aetiology and development of co-morbidity was important as it might provide the basis to understanding whether different disorders are actually distinct disorders or different expressions of an 'underlying disturbance'. Others have considered the possibility that co-morbidities are just undifferentiated accumulations of distress (Lilienfeld, 2003). The findings of the current study support these ideas regarding co-morbidity by suggesting that co-morbidities might just be different expressions of *p*, that occur simultaneously and sequentially during the life-course of individuals with higher risk.

Identification of a general psychopathology risk, if replicated and studied further, has the potential to unveil more distinct characteristics of specific disorders, alongside identifying disorder specific risk factors. This has implications for both the current nosologies in psychopathology, DSM and ICD, and for research, clinical practice and treatment models as discussed below.

The existence of the general psychopathology dimension that represents commonalities across various disorders suggests that future research into the aetiological factors, biological markers, environmental risk factors and expression of psychopathology will benefit from pooling together resources and having a more unified approach to studying *general psychopathology*, replacing or complementing present practice where disorders are mainly studied exclusively. This could increase our efficiency in the attempts towards the identification and understanding of the different factors and their interplay that result in psychiatric disorder. There is already much evidence that supports this route. The most compelling data can be found in the consistent empirical observation that environmental and demographic risk factors associated with most types of diagnosis/disorders are similar rather than disorder specific. This is further supported by recent genetic shared risk models from twin studies (Lahey et al., 2011) and studies of gene loci associations of psychopathology (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013) indicating that the gene correlates of major psychiatric disorders are more similar than heterogeneous. Additionally, partitioning out and studying disorders in their purer forms after accounting for *p* could enhance the identification of biomarkers and correlates that differentiate one disorder from another if these are indeed unique. Hence, as briefly illustrated by the current study, studying *p* not only will allow us to understand a general propensity for psychopathology, but then after partitioning for this factor can permit more detailed investigations of 'purer' forms of disorder to identify disorder specific cause/consequences.

Mental health treatment, for the most part has a different approach from most physical health treatment in that diagnosis and treatment are largely dependent on symptoms exhibited rather than the root or cause of the disorder. To illustrate the potential usefulness of further exploring this *general psychopathology* deficit, especially in terms of clinical utility and treatment, parallels can be drawn between a general psychopathology deficiency and immune-deficiencies in physical health. Immuno-deficiencies can be either hereditary or acquired; and immune-compromised individuals are more susceptible to infections. The type and nature of the diseases/infections that individuals with these deficiencies get vary greatly and depend upon several factors. Identifying the immunodeficiency rather than just observing expressed symptoms is crucial to long-term management and treatment in individuals with these immune deficits.

The existence of a propensity for general psychopathology suggests that individuals with greater risk or higher *p* are more likely to experience psychopathology irrespective of the form it might take during their lifetime, with environmental factors and life events only serving as moderators and triggers of the expression of specific disorders. If this is the case, as has been suggested (Caspi et al., 2013), individuals with higher '*p*' would be expected to transition through different diagnostic categories throughout their lifetime (Caspi et al., 2013). There is already some evidence that this is the case with disorders, both in childhood and adulthood (Angold et al., 1999; Moffitt et al., 2007). However, our understanding of the biological (genetic and neurological) correlates is only just emerging and explorations of gene-environment interactions and their role in psychopathology is in the very early stages. Judging by the many decades it took from discovering there is a immune deficit and then understanding its cell-level and genetic correlates (Rosen, 2000) we can only estimate that even with the increased scientific capacity we now have, there is a long way to go before our understanding of a general psychopathology liability factor can translate into realistic models of disease and its treatment.

General Discussion

The current thesis is structured in two parts. Part 1 regarding further validating a self-report mental health measure and Part 2 exploring the development and structure of internalising and externalising symptoms in two community based longitudinal cohorts spanning late childhood and early adolescence. The two main aims of Part 2 were: first, to examine the complexity of the development of self-reported symptoms of psychopathology in two longitudinal community based cohorts spanning late childhood and early adolescence, investigate the correlates of differential symptom development and their impact on another key domain of child functioning: academic attainment. Second, explore the possibility of an alternative structure of child psychopathology.

In this general discussion I first summarise the key findings of the six empirical studies. This is followed by a discussion of the overall strengths and limitations of the thesis. The next section outlines key themes that derive from the work aiming the draw out the wider implications of the studies and their findings ending in an exploration of future directions.

Summary of findings

The thesis has two main parts: examining self-reported measurement of symptoms in Part 1 (Studies 1 & 2) and exploring patterns of development, their correlates, impact and underlying structure of internalising and externalising symptoms in Part 2 (Studies 3 to 6).

The first two studies assessed the clinical validity and survey format equivalence of a child self-report measure of mental health, the M&MS questionnaire, which assesses symptoms in the two key domains of child psychopathology: internalising and externalising. In study 1, the clinical validity of the measure was established and results indicate that the M&MS discriminates between clinical and community samples to the same extent as other widely used measures such as the SDQ. Study 2 examined the psychometric equivalence of the paper and computer formats of the M&MS, bearing in mind the 'digital natives' status of current children and adolescents. The results demonstrate that the scale structure and

reliability are not different across formats, neither do the items operate differently in the two media. However, lower scale level scores are observed suggesting an overall dampening of scores with the paper-based survey, possibly representing higher levels of disclosure by young people on computers.

The next two studies investigated the complexity of individuals' symptom development by examining the developmental trajectories of internalising and externalising symptoms using latent class growth analysis, a method that allows the estimation of different person-centred trajectories of symptom development. Subsequently, the socio-demographic correlates, including gender, socio-economic status, ethnicity and age, of individuals with different trajectories were examined. This was followed by an assessment of the impact of these different types of trajectories on academic attainment. The externalising symptom trajectories in Study 3 demonstrated discernable patterns of symptom development in the short term that could be situated within the well-developed taxonomy of externalising behaviours (Moffitt, 1993). With regards to the impact on academic attainment, although this is a fairly well established negative longitudinal relationship, the study allowed a greater breakdown of the impact of differential symptom development on attainment and highlights clearly the negative consequences of developing externalising symptoms, especially in late childhood. In Study 4, trajectories of internalising symptoms, a much less studied area, were identified and correlates examined. The analysis revealed expected predictors such as gender and deprivation where, for instance, being female increased probability of stable or increasing trajectories in most cases with the notable exception of more males having a steep low-very high symptom trajectory during early adolescence, although overall numbers were not large in this group. The unexpected risk factor for higher symptom trajectories in the primary school sample was relative age in cohort, which significantly increased the probability of having higher symptoms in this age group. In terms of the impact of internalising symptoms on change in educational attainment from one key stage to the next the current study provides

some conclusive evidence that increasing internalising symptom trajectories are associated negatively with subsequent educational attainment in adolescence. Results for both internalising and externalising symptoms indicate that identifying greater heterogeneity in development allows for identification to some extent of the correlates or risk factors of different symptom development. In both studies, a trajectory based approach provided nuanced information about the impact of symptom development on academic attainment. In terms of the effect size, externalising symptom trajectories had larger negative impact compared to the internalising trajectories, but in both cases increasing or high symptom trajectories had a significant negative impact on educational attainment.

The fifth study focussed on the co-development of symptoms in the internalising and externalising domains in late childhood and early adolescence and aimed at uncovering patterns in their association and development over time in these two developmental periods. A striking finding was that not only are the two dimensions associated moderately at baseline, but their development over time is also moderately associated. More detailed analysis indicates that in the younger cohort, internalising difficulties at baseline are not significantly associated with development of externalising symptoms, whereas externalising symptoms at baseline significantly predict increasing internalising symptoms over time. In the older cohort both internalising and externalising symptoms at baseline predicted development of symptoms in the other domain over the three waves.

The consistency in the correlates and impact of these distinctive disorder types and the moderate degree of co-morbidity, both in terms of co-occurrence and co-development of symptoms, led to questions regarding the distinctiveness of these two dimensions widely studied as distinct dimensions in child psychopathology. Hence, the next and last study aimed to explore the dimensions themselves and investigated the underlying structure of child psychopathology. Using hierarchical bi-factor analysis, the study explored the possibility of a general risk factor for psychopathology. The general factor, along with the externalising and

internalising specific factors, fit the data well and demonstrated a model that can also help identify unique correlates of externalising and internalising dimensions. For example, the study indicates that once general risk of psychopathology is taken into account, deprivation is uniquely linked to externalising behaviours but not to internalising symptoms. Additionally the better predictive capacity of the general factor clearly demonstrates the promising potential of further studying the general factor, its biological and environmental correlates and could lead to better models of management and treatment of psychopathology.

Strengths and limitations

The studies in the thesis are strengthened by utilising two longitudinal datasets in different age cohorts that are broadly representative of the wider population at the respective ages. This is complemented by the various, more advanced methodologies used to examine the different research questions of interest. These two points in concert contribute to the strength of the thesis, by providing an opportunity to apply the advanced methodologies in data with sufficient power to allow the appropriate use of these methods.

As discussed in the first part of the thesis, evidence increasingly suggests children are able reporters of their health status, subject to being asked in an age-appropriate manner (Arseneault et al., 2005; Sharp et al., 2006). The systematic review of existing child self-report measures clearly highlighted a lack of a broad mental health measure that was suitable to use for primary school aged children, especially in a community context. Hence the further validation of a recently developed measure that was widely used as part of a national study, the Me and My School questionnaire, was undertaken. The current studies add to already existing validation and illustrates that the measure is psychometrically reliable and valid. Moreover this measure, unlike many validated child mental health measures (e.g., ASEBA, CHQ) was developed to be free to use which is crucial in reducing financial burden and administrative costs for large population based studies and routine use in schools and other community settings, making wider spread screening and assessment more accessible.

In spite of the evidence, in much research in the field of developmental psychopathology proxy reported symptoms, usually parent or teacher, are utilised to understand the development and structure of psychopathology. In the current thesis, focussing on using child self-reported symptoms, not only demonstrates that increased understanding, but also indicates that insights that might not be available from parent reported data, can be gained.

However, the use of *only* child reported symptoms due to a lack of proxy reported symptom data that could be analysed must still be acknowledged as a drawback. Although, as discussed above, child reports are important in gaining the individuals perspective, multiple reporters and triangulation are considered the best standard in developmental psychopathology research (De Los Reyes, 2013), owing to the high amount of disagreement between children and other reporters of their mental health (Achenbach et al., 1987, see Part 1 for more detailed discussion). Analysing data from other reporters, especially when exploring the general psychopathology factor in Study 6, might have contributed to greater understanding of the general psychopathology risk factor. However, as most studies use proxy reported symptoms to study development of symptoms, the current research works some way to redress the imbalance that is present in the current literature that is based mainly on other reported symptoms.

In terms of correlates of symptom development and the general and specific factors of psychopathology, some key socio-demographic correlates were included in analysis. However, the studies were limited in the number and type of risk factors and correlates that could be included, which stems from the existence of a limited number of contextual and environmental risk factors with known associations with development and psychopathology that could be collected as part of the large school based study whose data are used in the thesis.

Themes and implications

The implications of the findings of each of the separate studies are outlined in the discussion of the studies throughout the thesis. In this section I summarise wider themes present in this work and draw out the areas where the methods and results of the studies have potential relevance.

Individual focussed developmental psychopathology

One of the threads that runs through much of this work relates to emphasising the focus on the individual. The stress is on locating the individual, here the young person, at the centre of developmental psychopathology research. In the course of this thesis, this is represented in many different ways, including - focussing on the child as an individual with a voice who can be an able reporter of their health, the use of person-centered rather than variable centered methodologies to studying change and the discussions around an individual's propensity towards general psychopathology rather than the diagnostic category they fall within.

Self-reports of mental health

The use of the child's report of their own symptoms throughout this thesis has already been discussed above as a strength of the current work. In the background to Part 1, I laid out the current issues regarding measurement of mental health in the community setting, especially focussing on self-report measurement by young people. There has been much stress on proxy reported symptoms in developmental psychopathology, which to some extent rests on the belief that children and adolescents are unable to report on their own health sufficiently well. This belief is increasingly being challenged by the findings that suggest that when asked in an age and developmental stage appropriate manner young people are able reporters of their mental state (Arseneault et al., 2005; Sharp et al., 2006).

Although the thesis illustrates the measurement validity and usefulness in utilising child self-reports of health in developmental psychopathology research, the in-equivalence between the paper and computer formats of this measure raise some important issues pertaining to the digitalisation of data collection and ought to be discussed more by researchers using child reported measures in the various different survey formats currently available. Until we move into a future era where all data are collected electronically, and while presently this transition takes place, format based differences have the potential to impact on prevalence estimates, efficacy trials and direct comparisons of data. The findings regarding the differences in reporting based on survey format also strongly point to the importance of considering how we ask children regarding their mental well-being. Not just in terms of the wording and language used, as discussed in the introduction to Part 1, but also as indicated by results of Study 2 - the medium and setting. The results suggest that children might be more comfortable disclosing information online when compared to a paper-pencil survey. This raises thought provoking issues related to many aspects of how these data are collected. For instance, in collecting routine outcome measures in clinical practice, are there steps that can be taken to encourage CYP to feel comfortable, trust their setting and the person administering the measure and feel empowered to honestly report on their feelings. In school based studies, correspondingly, questions can be asked about the value placed on peers and teachers not finding out what is disclosed, and if there are ways in which confidentiality and privacy can be encouraged by the methods used in data collection. Hence, it would be interesting to explore the psychometric equivalence of children self-reporting mental health symptoms using different widely used measures of child mental health such as SDQ, YSR etc. as the implications of survey format differences have relevance not just for the M&MS but for child self-reported measures more generally.

Another key question that is raised by this debate surrounding self-reports is: why is self-reporting of symptoms considered an issue only for young people and not for adults?

What suddenly changes at age 18 that makes over 18 year olds appropriate reporters of their own mental health? There is little investigation into whether the reliability of children's reports of their mental health improves with age, which is an issue worth exploring as it is relevant to the sudden increase in trust in self-reporting around age 18 years.

These questions are linked with debates in health regarding patient involvement in decision making and how that applies to children, and an example can be seen in the controversy surrounding recent legislation allowing children euthanasia in Belgium. In light of this debate and legislations and policy such as "UN Convention on the Rights of the Child" 1989), Every Child Matters (Department for Education, 2003) and more recently government policy in the UK encouraging routine outcomes being monitored from the child's perspective as well (Department of Health, 2010). The ability of children to self-report on their health and well-being when asked appropriately also raises interesting possibilities about their involvement in other aspects of mental health service design and delivery including designing services and planning intervention.

Most relevantly, all these points highlight the importance of involving the individual or taking advice from young people in these decisions, especially as they become integral aspects of clinical care and school-based screening. Therefore, child self-report needs to not only stop being utilised in the minority in child psychopathology research but also increasingly recognised as the sole opportunity to examine development and mechanisms using measures that capture the individual's perspective.

At the same time it is important to note that the low associations between various reporters of children's mental health still needs further exploration (De Los Reyes, 2013). Whether and to what extent age/developmental period, characteristics of the child, characteristics of the reporters, context in which health status is recorded all contribute to the disparity in assessments is still worth understanding to help understand the properties of these

measurements better. Little is known about characteristics, developmental stage or even type of mental health difficulties being experienced that might affect the accuracy of self-reports and this topic warrants further investigation.

In terms of implications for practice, the literature abounds with evidence in support of early intervention: the financial costs of difficulties and consequently the financial savings of intervening early (Suhrcke, Pillas, & Selai, 2008), the efficacy of early interventions (Adi, Kiloran, Janmohamed, & Stewart-Brown, 2007), the benefit for the functioning and outcomes of the individual and also their family, classrooms and wider society (Greenberg et al., 2003). The further validation of the M&MS questionnaire which is brief, easy to read and free to use, supports the possibility of more universal screening strategies in community settings, especially schools.

Developmental complexity

"*The study of development is foremost the study of change*" (Hartman, Pelzel, & Abbott, 2011, pp.117), making studying changes in symptoms one of the key interests of developmental psychopathology. The importance of studying complexities in symptom development has been stressed through the decades (Cicchetti & Rogosch, 2002; Sroufe & Rutter, 1984), and becomes more relevant to studying aetiology and risk of psychopathology as more longitudinal data is available to us. The current thesis demonstrates how it is also relevant to studying impact of adaptive and maladaptive symptom development on other domains and dissecting longitudinal relationships between domains of functioning: a key tenet of developmental psychopathology and ecosystems theories and approaches (Cicchetti & Toth, 2009; Lewis, 2000; Magnusson, 1995). The identification of independent growth trajectories that better represent the data demonstrates the relevance of understanding heterogeneity in the development of psychopathology. Studies of person centred trajectories have tended to focus on long term trajectories of symptoms through childhood and adolescence, and focus more on placing these in the life-course perspective of development.

The current studies indicate how shorter term ups and downs in symptoms can also be examined with possibly studying peer, school and other influences through crucial periods of late childhood and early adolescence. In the current thesis this was applied successfully to further clarify the impact of symptom development on academic attainment. These approaches to studying person-centred development also allow, to some extent, explorations of differential susceptibility to risk factors. However, as can be seen from studies 3 and 4, although some information is gained about broad correlates of differential symptom trajectories, the correlates are more similar than different: a finding that is consistent with many of the well-established correlates of child psychopathology (egs. Green et al., 2005; Varese et al., 2012). This finding draws attention to the limited information gained from the variable focussed methods employed in psychopathology research as different individuals are treated in terms of their descriptive characteristics rather than an individual with characteristics that interact with their genetic make-up, environment and their personal developmental history (Bergman et al., 2006).

The evidence from the current studies, that developing symptoms does impact negatively on functioning in other key areas of children's functioning such as learning has implications for school policy and practice. Moreover, in the younger sample even sub-threshold stable and decreasing from above threshold internalising symptoms were associated with worse later educational attainment suggesting a development-lag effect, where difficulties at a crucial developmental stage still had impact on future learning even if problems decreased or were at sub-clinical levels. These findings lend support to the arguments for the need for school based promotion and interventions to prevent mental health problems, which are especially relevant in light of increasing focus on academic outcomes and the relegation of pupils' well-being to the back seat (Greenberg, 2010; Shoshani & Steinmetz, 2013).

General psychopathology factor

Given the results of Study 6, the existence of a more general risk or psychopathology deficit posits interesting implications for research in developmental psychopathology. Not only does it present a more unified model of psychopathology as being a broad dimension, it also has the potential to assist in our attempts to find clearer correlates of specific disorders or domains such as internalising and externalising symptoms. If indeed a more inherent propensity to psychopathology does exist, it suggests that individuals with greater propensity are more likely to experience psychopathology no matter what, with environmental factors and life events serving as moderators and triggers of the expression of specific disorders. The general psychopathology factor situates the discussion about risk in individual people's differential vulnerability and resilience to psychopathology and the role of environmental triggers or stressors that lead to expressing clinical levels of psychopathology.

The ability to identify and describe a general psychopathology factor has the potential to contribute to the improving our understanding of varied aspects of development and psychopathology, including: 1) differential susceptibility of individuals to disorder, 2) the presence of comorbid symptoms, 3) occurrences of both homo-typic and hetero-typic continuity throughout the life-course, and 4) the observed increasing diagnostic specificity with age. The concept has relevance for treatment as well and might lead to better treatment models, where there is a more holistic focus on the individual instead of the presented diagnosis. This might also promote provision of intervention that is tailored to the individual, their risk factors, situation and need rather than solely by their diagnosis.

As discussed in the background to Part 2 previously, there have been attempts and suggestions to try and integrate the various problems in studying psychopathology such as co-morbidity, dimensionality vs. categories/diagnosis and explanations of heterotypic and homo-typic continuities of symptoms through the life-course. The general psychopathology factor or propensity has the potential to be the structural umbrella under which these concepts can be clarified and better understood. For instance a general propensity towards psychopathology,

places the emphasis on considering the individual's inherent propensity for psychopathology rather than just focussing on dimensions and diagnostic categories and symptoms present. The understanding then would involve evaluating any particular individuals' propensity alongside the risk factors that might have triggered a particular set of symptoms. This would take into consideration a person's developmental history, previous psychopathology and the specific current manifestation of symptoms, an approach which is consistent with the life-course perspectives that are becoming more widely accepted and utilised in developmental research (Maughan & Rutter, 2009). In terms of co-morbidity, concurrent presentations of disorder can be reformulated as an expression of psychopathology being expressed in varied ways simultaneously.

There is much discussion in the literature regarding the limited utility of diagnostic categories in researching development of psychopathology (Jensen et al., 2006). However, there is a greater amount of focus on the different manifestations of psychopathology that form the nosologies such as the DSM and ICD. These classifying systems have their many uses in terms of exploring what therapies work for what types of conditions, acting as an easy heuristic tool when delivering treatment. An unintended adverse consequence results in a system where the focus is sometimes too much on the condition, and not enough on the individual. Additionally bulks of evidence indicate that the efficacy of treatments is explained to a large extent by elements of treatment that are non-specific to particular treatment approach (Norcross & Wampold, 2011; Wampold, 2013). If that is indeed the case the argument for focussing on the individual and their developmental propensity, environmental risk and personal history is further supported in terms of treatment and care as well.

The existence of a general psychopathology deficit that might increase or decrease the propensity of individuals shifts the focus from diagnostic categories to a 'psycho-immuno-deficiency'. The utility of such an approach will become clearer as more investigation is conducted into this general factor. This has the potential not only to obtain better theoretical

models of psychopathology but also has potential utility in a clinical setting, bearing in mind that more than half of young people receiving mental health support manifest pathology across several different diagnostic categories (Martin & de Francesco, 2014).

Considering the substantive change the concept of general psychopathology risk and the various areas of psychopathology research and understanding this has the potential to affect, it is worth recognising that these developments contribute to a shifting paradigm within which we study psychopathology and development. However much more research is necessary to develop both the theoretical models and the practical application in this area, especially as it challenges currently established ways of conceptualising and studying psychopathology. The many potential areas of further investigation are discussed below in future directions.

Hence, in conclusion to this broader theme of individual focussed research, the increased orientation on the individual's symptoms, development and characteristics start to situate the investigation in theories of differential susceptibility to risk (Belsky et al., 2007), multi- and equi-finality of markers and outcomes (Cicchetti & Rogosch, 1996) and places the emphasis on the individual's cumulative developmental history, placing the individual's development at the center of the system - as it determines in part the person's developmental environment, their genetic expressions and the interactions between the multiple levels and factors (Sroufe, 2007).

The next theme briefly discusses the gains to be made by sometimes- where necessary, using methodologies that might help uncover more nuances in the data.

Old questions, new methods

All through this work, I have made an attempt to think about the limitations of the more commonly used approaches, in many cases the accepted ways of analysing these types of data to allow myself to consider the alternatives. The methodology literature always

offered up many alternatives: some simple, some complex. However, although many of the methods used in the current studies have existed from a few years to a few decades, often they were not being employed in the study of child psychopathology. For instance, in the measurement studies I utilised propensity score matching to try to ensure that up to the extent possible using this method, the samples in the different conditions were similar on key demographic variables that we know are associated with mental health. The more common practice of ensuring no differences overall between groups on key variables is limited by aggregates, which can lead to differences between groups being due to variations in distribution between the two groups on key variables (Dehejia & Wahba, 2002). Another example is the adaptation of DIF analysis from comparing socio-demographic groupings to comparing survey format effects. In both these cases the method has been available to us for many years; however, they have not been applied widely in the field to answer these sorts of questions.

The identification of independent growth trajectories that better represent the data demonstrates the relevance of understanding heterogeneity in the development of psychopathology. Methodologically, these studies illustrate a person-centered approach to studying longitudinal impact of development in one domain on development in another domain, an approach that could be developed further and applied to studying nuances in the longitudinal impacts of development in other domains and investigating other longitudinal relationships.

Another example is the analysis in study 5, which involved mapping developmental trajectories in internalising and externalising domains onto one another leading to some interesting insights into the overlap between trajectories in these two domains. However, the descriptive nature of the analysis and the lack of further detailed exploration in terms of correlates, highlighted the lack of available methods to allow more detailed study of correlates and predictors of co-occurring symptom types using such an approach.

Longitudinal change and processes are complex, multi-level and multi-faceted, for instance, studying changes in one outcome variable over time is not simple as change may be gradual and linear or can be sudden and non-linear, and whether this is captured or not depends on many factors including measurement tools and timing, making it very easy to miss apt time periods or moments to study them (Hartman et al., 2011). In this process, investigating other covariates of interest, which also in many cases are dynamic and changing highlights the challenges faced by developmental researchers in understanding processes of change. However, developments in methodology and the availability of more longitudinal data is ever increasing the possibilities of what can be investigated.

It is important to note that all methods have their limitations and when using one set of methods that might help uncover certain associations, there is sometimes a risk that some other aspect of the analysis is compromised. For instance, the trajectories provide a more heterogeneous representation of symptom development at a group level but at the same time risks considering all individuals within the same trajectory group as being homogeneous. Hence, using multiple approaches to answer questions might lead to understanding better different aspects of the question of interest. These different approaches and nuances of the relationships of interest could, like different pieces of a puzzle, taken together, start to give a clearer understanding of these complex developmental relationships.

Future directions

Many interesting questions and topics for future research arise from the results of the various studies reported in this thesis. Some example topics for further investigation include, for instance, survey format effects in child mental health and examining the effect of relative age in cohort on mental health outcomes. However, the current section does not aim to identify and list them all. Instead I will discuss the potential interesting directions that follow from the last study of the thesis- a general propensity for psychopathology.

The current thesis lays the foundation blocks for the identification and initial establishment of the conceptual and predictive validity of the general psychopathology factor in young people. The potential relevance and importance of studying general risk or deficit for psychopathology is outlined in detail Study 6, and has the potential to aid research and understanding in several areas of study as outlined above in the implications section.

Given the very early days of the investigation into this general risk for psychopathology, the topic warrants a complete programme of investigation to first establish the construct, understand its correlates and predictive utility and also investigate how it might be measured. Research with different datasets, using different measures, different reporters and exploring various diverse correlates is necessary to further understand this general risk factor and begin the process of investigating it's measurement, identification and informing treatment models. Studies might aim to understand the genetic and shared environment influences on the general psychopathology factor using twin samples. Neurological correlates can be explored in brain imaging studies and genome wide association studies might help illuminate the genetic correlates of mental disorder. The creation of a measure or a method to directly measure this 'general psychopathology factor' would also be a necessary step if future research and practice is to be better able to focus on this variable.

In summary, the hypothesis of a general psychopathology factor has the potential to have a real impact on the field of developmental psychopathology and allied disciplines as outlined above, however, more research is required before the potential usefulness of the construct, both theoretically and practically, becomes clearer.

Conclusion

The thesis contributes to accumulating evidence to support greater understanding of the complexities in symptom development in young people. At the same time it explores the possibility of a simpler structural model to help describe, understand and treat psychopathology. The result is a thesis where the stress is on individual focussed developmental psychopathology research, all the while underscoring the value of child self-reports of symptoms and modern data analytic techniques in studying the structure, development and cross-domain impact of child psychopathology.

References

Achenbach, T. M. (1966). The Classification Of Children's Psychiatric Symptoms: A Factor-Analytic Study. *Psychological Monographs: General and Applied, 80*(7), 1-37.

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4–18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: a review and analysis of empirical efforts. *Psychological Bulletin, 85*(6), 1275.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles.* Burlington, VT: University of Vermont Research Center for Children, Youth and Families.

Adams, S. J. (2002). Educational Attainment and Health: Evidence from a Sample of Older Adults. *Education Economics, 10*(1), 97-109.

Adi, Y., Kiloran, A., Janmohamed, K., & Stewart-Brown, S. (2007). Systematic review of the effectiveness of interventions to promote mental wellbeing in primary schools Report 1: Universal approaches which do not focus on violence or bullying *National Institute for Health and Clinical Excellence, London*. Retrieved from: http://www.nice.org.uk/nicemedia/pdf/MentalWellbeingWarwickUniReview1.pdf.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). Standards for educational and psychological testing. Washington, DC: American Psychological Association.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). N.J: Upper Saddle River.

Anderson, J. C., Williams, S., McGee, R., & Silva, P. A. (1987). DSM-III disorders in preadolescent children: Prevalence in a large sample from the general population. *Archives of general psychiatry, 44*(1), 69.

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal Of Child Psychology And Psychiatry, 40*(1), 57-87.

Angold, A., Costello, E. J., & Worthman, C. M. (1998). Puberty and depression: the roles of age, pubertal status and pubertal timing. *Psychological Medicine, 28*(1), 51-61.

Angold, A., Erkanli, A., Farmer, E. M. Z., Fairbank, J., Burns, B., Keeler, G., & Costello, E. J. (2002). Psychiatric disorder, impairment, and service use in rural African American and white youth. *Archives of general psychiatry, 59*(10), 893.

Ansary, N. S., & Luthar, S. S. (2009). Distress and academic achievement among adolescents of affluence: A study of externalizing and internalizing problem behaviors and school performance. *Development and Psychopathology, 21*(01), 319-341.

Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Psychometric evaluation of 5-and 7-year-old children's self-reports of conduct problems. *Journal of Abnormal Child Psychology, 33*(5), 537-550.

Aseltine, R. H., Gore, S., & Colten, M. E. (1998). The co-occurrence of depression and substance abuse in late adolescence. *Development and Psychopathology, 10*(03), 549-570.

Bardone, A. M., Moffitt, T. E., Caspi, A., Dickson, N., & Silva, P. A. (1996). Adult mental health and social outcomes of adolescent girls with depression and conduct disorder. *Development and Psychopathology, 8*(04), 811-829.

Barker, E., & Maughan, B. (2009). Differentiating early-onset persistent versus childhood-limited conduct problem youth. *American Journal of Psychiatry, 166*(8), 900-908.

Barro, R. J., & Lee, J.-W. (2001). International data on educational attainment: updates and implications. *Oxford Economic Papers, 53*(3), 541-563.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*(4), 757-786.

Baum, A., Garofalo, J., & Yali, A. (1999). Socioeconomic status and chronic stress: Does stress account for SES effects on health? *Annals of the New York Academy of Sciences, 896*(1), 131-144.

Beck, J., Beck, A., & Jolly, J. (2001). *Beck Youth Inventories of Emotional & Social Impairment manual*. San Antonio, TX Psychological Corporation.

Belsky, J., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2007). For Better and For Worse: Differential Susceptibility to Environmental Influences. *Current Directions in Psychological Science, 16*(6), 300-304.

Bergman, L. R., von Eye, A., & Magnusson, D. (2006). Person-oriented research strategies in developmental psychopathology. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology, Volume 1: Theory and Methods*. NJ: Wiley.

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, 47-53.

Beyers, J. M., & Loeber, R. (2003). Untangling Developmental Relations Between Depressed Mood and Delinquency in Male Adolescents. *Journal of Abnormal Child Psychology, 31*(3), 247-266.

Bland, J. M., & Altman, D. G. (1994). Matching. *British Medical Journal, 309*, 1128.

Boardman, M. (2006). The impact of age and gender on prep children academic achievements. *Australian Journal of Early Childhood, 31*(4), 1-6.

Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior, 23*(1), 463-477.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*(3), 345-370.

Broeren, S., Muris, P., Diamantopoulou, S., & Baker, J. (2013). The Course of Childhood Anxiety Symptoms: Developmental Trajectories and Child-Related Factors in Normal Children. *Journal of Abnormal Child Psychology, 41*(1), 81-95.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. a., . . . Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39*, 222-245.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Eds.), *Handbook of Child Psychology:Theoretical models of human development Wiley* (6th ed., pp. 793–828). Hoboken, NJ: Wiley.

Brooks, T. L., Harris, S. K., Thrall, J. S., & Woods, E. R. (2002). Association of adolescent risk behaviors with mental health symptoms in high school students. *Journal of Adolescent Health, 31*(3), 240-246.

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A Tutorial on Hierarchically Structured Constructs. *Journal of Personality, 80*(4), 796-846.

Bucholz, K., Hesselbrock, V., Heath, A., Kramer, J., & Schuckit, M. (2000). A latent class analysis of antisocial personality disorder symptom data from a multi-centre family study of alcoholism. *Addiction, 95*(4), 553-567.

Bullinger, M., & Ravens-Sieberer, U. (1995). Health related QOL assessment in children: A review of the literature. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée, 45*, 245-254.

Burns, B. J., Costello, E. J., Angold, A., Tweed, Stangl, D., Farmer, E. M. Z., & Erkanli, A. (1995). Children's Mental Health Service Use Across Service Sectors. *Health Affairs, 14*(3), 147-159.

Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, and discriminant validity. *American Psychologist, 15*(8), 546-553.

Cantwell, D. P., Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1997). Correspondence Between Adolescent Report and Parent Report of Psychiatric Diagnostic Data. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(5), 610-619.

Capaldi, D. M. (1991). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: I. Familial factors and general adjustment at Grade 6. *Development and Psychopathology, 3*(03), 277-300.

Capaldi, D. M. (1992). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: II. A 2-year follow-up at Grade 8. *Development and Psychopathology, 4*(01), 125-144.

Capaldi, D. M., & Stoolmiller, M. (1999). Co-occurrence of conduct problems and depressive symptoms in early adolescent boys: III. Prediction to young-adult adjustment. *Development and Psychopathology, 11*(01), 59-84.

Caron, C., & Rutter, M. (1991). Comorbidity in Child Psychopathology: Concepts, Issues and Research Strategies. *Journal Of Child Psychology And Psychiatry, 32*(7), 1063-1080.

Carter, A. S., Godoy, L., Wagmiller, R. L., Veliz, P., Marakovitz, S., & Briggs-Gowan, M. J. (2010). Internalizing trajectories in young boys and girls: The whole is not a simple sum of its parts. *Journal of Abnormal Child Psychology, 38*(1), 19-31. Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., . . . Moffitt, T. E. (2013). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*.

Cerdá, M., Sagdeo, A., & Galea, S. (2008). Comorbid Forms of Psychopathology: Key Patterns and Future Research Directions. *Epidemiologic Reviews, 30*(1), 155-177.

Chen, H., Cohen, P., & Chen, S. (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation, 39*(4), 860-864.

Chen, X., Huang, X., Chang, L., Wang, L., & Li, D. (2010). Aggression, social competence, and academic achievement in Chinese children: A 5-year longitudinal study. *Development and Psychopathology, 22*(03), 583-592.

Children Act, (2004).

Chorpita, B. F., Reise, S., Weisz, J. R., Grubbs, K., Becker, K. D., & Krull, J. L. (2010). Evaluation of the Brief Problem Checklist: child and caregiver interviews to measure clinical progress. *Journal of Consulting and Clinical Psychology, 78*(4), 526.

Chorpita, B. F., Yim, L., Moffitt, C., Umemoto, L. A., & Francis, S. E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. *Behaviour Research and Therapy, 38*(8), 835-855.

Cicchetti, D. (1984). The Emergence of Developmental Psychopathology. *Child Development, 55*(1), 1-7.

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology, 8*(04), 597-600.

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of consulting and clinical psychology, 70*(1), 6.

Cicchetti, D., & Toth, S. L. (2009). The past achievements and future promises of developmental psychopathology: the coming of age of a discipline. *Journal Of Child Psychology And Psychiatry, 50*(1-2), 16-25.

Clark, L. A. (2005). Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology, 114*(4), 505.

Clark, S. L., & Muthén, B. O. (2009). Relating latent class analysis results to variables not included in the analysis. http://statmodel2.com/download/relatinglca.pdf.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (second ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, D. A. (1990). Relation of social and academic competence to depressive symptoms in childhood. *Journal of Abnormal Psychology, 99*, 422-429.

Cole, D. A., Martin, J. M., Powers, B., & Truglio, R. (1996). Modeling Causal Relations Between Academic and Social Competence and Depression: A Multitrait-Multimethod Longitudinal Study of Children. *Journal of Abnormal Psychology, 105*(2), 258-270.

Copeland, W. E., Shanahan, L., Costello, E. J., & Angold, A. (2009). Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Archives of general psychiatry, 66*(7), 764-772.

Costello, J. (2009). Using epidemiological and longitudinal approaches to study causal hypothesis. In M. e. a. Rutter (Ed.), *Rutter's Child and Adolescent Psychiatry* (pp. 58-70). Oxford: Blackwell Publishing Ltd.

Costello, E. J., Compton, S. N., Keeler, G., & Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *JAMA, 290*(15), 2023-2029.

Costello, E. J., Copeland, W., & Angold, A. (2011). Trends in psychopathology across the adolescent years: What changes when children become adolescents, and when adolescents become adults? *Journal of child psychology and psychiatry, and allied disciplines, 10*, 1015-1025.

Costello, E. J., Erkanli, A., Copeland, W., & Angold, A. (2010). Association of family income supplements in adolescence with development of psychiatric and substance

use disorders in adulthood among an american indian population. *JAMA, 303*(19), 1954-1960.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of general psychiatry, 60*, 837-844.

Côté, S. M., Boivin, M., Liu, X., Nagin, D. S., Zoccolillo, M., & Tremblay, R. E. (2009). Depression and anxiety symptoms: onset, developmental course and risk factors during early childhood. *Journal Of Child Psychology And Psychiatry, 50*(10), 1201-1208.

Coyne, J. C. (1976). Depression and the response of others. *Journal of Abnormal Psychology, 85*(2), 186.

Cremeens, J., Eiser, C., & Blades, M. (2006). Factors influencing agreement between child self-report and parent proxy-reports on the Pediatric Quality of Life Inventory 4.0 (PedsQL) generic core scales. *Health Qual Life Outcomes, 4*, 58.

Crocetti, E., Klimstra, T., Keijsers, L., Hale, W., III, & Meeus, W. (2009). Anxiety trajectories and identity development in adolescence: a five-wave longitudinal study. *Journal of Youth and Adolescence, 38*(6), 839-849.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet, 381*, 1371–1379.

Crystal, S., Shea, D., & Krishnaswami, S. (1992). Educational attainment, occupational history, and stratification: determinants of later-life economic outcomes. *Journal of Gerontology, 47*(5), 213-221.

Cummings, E. M., & Cicchetti, D. (1993). Toward a Transactional Model of Relations between Attachment and Depression. In M. T. Greenberg, D. Cicchetti & E. M. Cummings (Eds.), *Attachment in the preschool years: Theory, research, and intervention* (pp. 339-372). Chicago: University of Chicago Press.

Curran, P. J., & Hussong, A. M. (2003). The use of latent trajectory models in psychopathology research. *Journal of Abnormal Psychology, 112*(4), 526.

De Los Reyes, A. (2013). Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Development and Psychopathology, 25*(03), 669-682.

De Los Reyes, 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*(4), 483-509.

De Pauw, S. S., & Mervielde, I. (2010). Temperament, personality and developmental psychopathology: A review based on the conceptual dimensions underlying childhood traits. *Child Psychiatry & Human Development, 41*(3), 313-329.

Dehejia, R. H., & Wahba, S. (2002). Propensity Score-Matching Methods For Nonexperimental Causal Studies. *The Review of Economics and Statistics, 84*(1), 151-161.

Deighton, J., Patalay, P., Belsky, J., Humphrey, N., Vostanis, P., Fugard, A., . . . Wolpert, M. (2013). Targeted mental health provision for children with behaviour difficulties in primary schools: Results of a national randomised trial *The Psychology of Education Review, 37*(2), 40-47.

Deighton, J., Tymms, P., Vostanis, P., Belsky, J., Fonagy, P., Brown, A., . . . Wolpert, M. (2013). The Development of a School-Based Measure of Child Mental Health. *Journal of Psychoeducational Assessment, 31*(3), 247-257.

Dekker, M. C., Ferdinand, R. F., Van Lang, N. D. J., Bongers, I. L., Van Der Ende, J., & Verhulst, F. C. (2007). Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome. *Journal Of Child Psychology And Psychiatry, 48*(7), 657-666.

Department for Education. (2003). Every child matters. London: Department for Education and Skills.

Department of Health. (2010). *Equity and Excellence: Liberating the NHS*. London: Crown.

DFE. (2010a). *Children with special educational needs 2010: An analysis*. London: DFE Publications.

DFE. (2010b). *Schools, pupils and their characteristics*. Retrieved from https://www.gov.uk/government/publications/schools-pupils-and-their-characteristics-january-2010.

DiLalla, L. F., Marcus, J. L., & Wright-Phillips, M. V. (2004). Longitudinal effects of preschool behavioral styles on early adolescent school performance. *Journal of School Psychology, 42*(5), 385-401.

Dishion, T. J., & Patterson, G. R. (2006). The development and ecology of antisocial behavior in children and adolescents. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology, Vol 3: Risk, disorder, and adaptation* (Vol. 3, pp. 503-541). Hoboken, NJ, US: John Wiley & Sons

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65*(2), 649-665.

Du Fort, G. G., Newman, S. C., & Bland, R. C. (1993). Psychiatric comorbidity and treatment seeking: Sources of selection bias in the study of clinical populations. *Journal of Nervous and Mental Disease, 181*(8), 467-474.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School Readiness and Later Achievement. . *Developmental Psychology 43*(6), 1428-1446.

Eccles, J. S., & Roeser, R. W. (2011). School and Community Influences on Human Development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental Science: An Advanced Textbook* (6th ed.). New York and Hove: Psychology Press.

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence, 21*(1), 225-241.

Eiser, C., & Kopel, S. (1997). Children's perceptions of health and illness. In K. J. Petrie & J. A. Weinman (Eds.), *Perceptions of health and illness : current research and applications.* Amsterdam: Harwood Academic.

Eiser, C., & Morse, R. (2001). A review of measures of quality of life for children with chronic illness. *Archives of Disease in Childhood, 84*, 205-211.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum.

Emerson, E. (2003). Prevalence of psychiatric disorders in children and adolescents with and without intellectual disability. *Journal of Intellectual Disability Research, 47*(1), 51-58.

Emmons, R. A. (1995). Levels and domains in personality: An introduction. *Journal of Personality, 63*(3), 341-364.

Epstein, M. H. (1999). The development and validation of a scale to assess the emotional and behavioral strengths of children and adolescents. *Remedial and Special Education, 20*, 258-262.

Epstein, M. H. (2004). *Behavioral and Emotional Rating Scale–Second edition*. Austin, TX:: PRO-ED.

Epstein, M. H., Harniss, M. K., Ryser, G., & Pearson, N. (1999). The Behavioral and Emotional Rating Scale: Test-retest and inter-rater reliability. *Journal of Child and Family Studies, 8*(319-327).

Fauber, R., Forehand, R., Long, N., Burke, M., & Faust, J. (1987). The relationship of young adolescent Children's Depression Inventory (CDI) scores to their social and cognitive functioning. *Journal of Psychopathology and Behavioral Assessment, 9*(2), 161-172.

Fergusson, D. M., Horwood, L. J., & Boden, J. M. (2006). Structure of internalising symptoms in early adulthood. *The British Journal of Psychiatry, 189*(6), 540-546.

Fergusson, D. M., Horwood, L. J., Ridder, E. M., & Beautrais, A. L. (2005). Subthreshold depression in adolescence and mental health outcomes in adulthood. *Archives of general psychiatry, 62*(1), 66.

Fergusson, D. M., & Woodward, L. J. (2002). Mental health, educational, and social role outcomes of adolescents with depression. *Archives of general psychiatry, 59*(3), 225-231.

Figlio, D. N. (2007). Boys Named Sue: Disruptive Children and Their Peers. *Education Finance and Policy, 2*(4), 376-394.

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Charlotte, NC: Information Age Publishing.

Flanagan, D. P., Alfonso, V. C., Primavera, L. H., Poval, L., & Higgins, D. (1996). Convergent validity of the BASC and SSRS: Implications for social skills assessment. *Psychology in the Schools, 33*, 13-23.

Flanagan, R. (1995). A review of the Behavior Assessment System for Children (BASC)Assessment consistent with the requirements of the Individuals with Disabilities Education Act (IDEA). *Journal of School Psychology, 33*(2), 177-186.

Fleming, J. E., Offord, D. R., & Boyle, M. H. (1989). Prevalence of childhood and adolescent depression in the community. Ontario Child Health Study. *The British Journal of Psychiatry, 155*(5), 647-654.

Fonagy, P., Twemlow, S. W., Vernberg, E. M., Nelson, J. M., Dill, E. J., Little, T. D., & Sargent, J. A. (2009). A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression. *Journal Of Child Psychology And Psychiatry, 50*(5), 607-616.

Ford, T., Goodman, R., & Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 1203-1211.

Ford, T., Hamilton, H., Goodman, R., & Meltzer, H. (2005). Service Contacts Among the Children Participating in the British Child and Adolescent Mental Health Surveys. *Child and Adolescent Mental Health, 10*(1), 2-9.

Fredriksson, P., & Öckert, B. (2005). Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance. *IZA Discussion Paper No. 1659*.

Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U. I., Marttunen, M. J., Koivisto, A.-M., & Kaltiala-Heino, R. (2008). Depression and school performance in middle adolescent boys and girls. *Journal of adolescence, 31*(4), 485-498.

Furnham, A. (1986). Response bias, social desirability and dissimulation. . *Personality and Individual Differences, 7*(3), 385-400.

Galambos, N. L., Barker, E. T., & Krahn, H. J. (2006). Depression, Self-Esteem, and Anger in Emerging Adulthood: Seven-Year Trajectories. [Article]. *Developmental Psychology, 42*(2), 350-365.

Gallagher, M. W., Lopez, S. J., & Preacher, K. J. (2009). The Hierarchical Structure of Well-Being. *Journal of Personality, 77*(4), 1025-1050.

Garnefski, N., & Diekstra, R. W. (1997). "Comorbidity" of behavioral, emotional, and cognitive problems in adolescence. *Journal of Youth and Adolescence, 26*(3), 321-338.

Gazelle, H., & Ladd, G. W. (2003). Anxious solitude and peer exclusion: A diathesis–stress model of internalizing trajectories in childhood. *Child Development, 74*(1), 257-278.

Gilman, S. E., Kawachi, I., Fitzmaurice, G. M., & Buka, S. L. (2002). Socioeconomic status in childhood and the lifetime risk of major depression. *International Journal of Epidemiology, 31*(2), 359-367.

Goldberg, D., & Goodyer, I. M. (2005). *The origins and course of common mental disorders*. New York: Routledge.

Goodman, A., Patel, V., & Leon, D. (2010). Why do British Indian children have an apparent mental health advantage? *Journal of child psychology and psychiatry, and allied disciplines, 51*, 1171-1183.

Goodman, A., Patel, V., & Leon, D. A. (2008). Child mental health differences amongst ethnic groups in Britain: A systematic review. *BMC Public Health, 8*(258-268).

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581-586.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1337-1345.

Goodman, R., Gledhill, J., & Ford, T. (2003). Child psychiatric disorder and relative age within school year: cross sectional survey of large population sample. *BMJ, 327*(7413), 472.

Goodman, R., Meltzer, H., & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. . *European Child and Adolescent Psychiatry, 7*, 125-130.

Gotlib, I. H., & Beatty, M. E. (1985). Negative responses to depression: The role of attributional style. *Cognitive Therapy and Research, 9*(1), 91-103.

Gotlib, I. H., & Meltzer, S. J. (1987). Depression and the perception of social skill in dyadic interaction. *Cognitive Therapy and Research, 11*(1), 41-53.

Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). Mental health of children and young people in Great Britain, 2004. In C. Copyright (Ed.). Basingstoke, Hampshire.

Green, J. G., McLaughlin, K. A., Berglund, P. A., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2010). Childhood adversities and adult psychiatric disorders in the national comorbidity survey replication I: associations with first onset of DSM-IV disorders. *Archives of general psychiatry, 67*(2), 113-123.

Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C.-P., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of Consulting and Clinical Psychology, 73*(2), 229.

Greenberg, M. T. (2010). School-based prevention: current status and future challenges. *Effective Education, 2*(1), 27-52.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American psychologist, 58*(6-7), 466-474.

Gregg, P., & Machin, S. (2000). The relationship between childhood experiences, subsequent educational attainment and adult labour market performance *Child Well Being in Modern Nations: What do we Know* (pp. Last accessed on: 25.July.2014): London School of Economics and Social Sciences.

Grimm, P. (2010). Social Desirability Bias *Wiley International Encyclopedia of Marketing*: John Wiley & Sons, Ltd.

Gumora, G., & Arsenio, W. F. (2002). Emotionality, emotion regulation, and school performance in middle school children. *Journal of School Psychology, 40*(5), 395-413.

Guralnik, J. M., Land, K. C., Blazer, D., Fillenbaum, G. G., & Branch, L. G. (1993). Educational status and active life expectancy among older blacks and whites. *New England Journal of Medicine, 329*(2), 110-116.

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis, 40*(2), 171-178.

Hale, D., Patalay, P., Fitzgerald-Yau, N., Hargreaves, D., Bond, L., Görzig, A., . . . Viner, R. (2013). School-Level Variation in Health Outcomes in Adolescence: Analysis of Three Longitudinal Studies in England. *Prevention Science*, 1-11.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of Classical Test Theory and Item Response Theory and their applications to test development ITEMS. Madison, WI:National Council on Measurement in Education.

Hammen, C., & Watkins, E. (2008). *Depression* (2nd ed.). Hove:East Sussex: Psychology Press.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., Mcgee, R., & Angell, K. E. (1998). Development of Depression From Preadolescence to Young Adulthood : Emerging Gender Differences in a 10-Year Longitudinal Study *Journal of Abnormal Psychology, 107,* 128-140.

Harden, K. P., & Mendle, J. (2012). Gene-environment interplay in the association between pubertal timing and delinquency in adolescent girls. *Journal of Abnormal Psychology, 121*(1), 73.

Hartman, D. P., Pelzel, K. E., & Abbott, C. B. (2011). Design, measurement and analysis in developmental research. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook (6th ed.)* (pp. 109-197). New York: Psychology Press.

Hayslett, M. M., & Wildemuth, B. M. (2004). Pixels or pencils? The relative effectiveness of Web-based versus paper surveys. *Information Science Research, 26*(1), 73-93.

Helsper, E., & Eynon, R. (2009). Digital natives: where is the evidence? . *British Educational Research Journal, 36*(3), 503-520.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms.*Psychological Bulletin, 111*(1), 127-155.

Hinshaw, S. P., & Anderson, C. A. (1996). Conduct and oppositional defiant disorders. In E. J. Marsh & R. A. Barkley (Eds.), *Child psychopathology*. New York: Guilford Press.

Hobbs, G., & Vignoles, A. (2010). Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal 36*(4), 673-690.

Holländare, F., Andersson, G., & Engström, I. (2010). A Comparison of Psychometric Properties Between Internet and Paper Versions of Two Depression Instruments (BDI-

II and MADRS-S) Administered to Clinic Patients. *Journal of Medical Internet Research, 12*(5).

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Humphrey, N., Lendrum, A., Barlow, A., Wigelsworth, M., & Squires, G. (2013). Achievement for All: Improving psychosocial outcomes for students with special educational needs and disabilities. *Research in Developmental Disabilities, 34*(4), 1210-1225.

Jackson, K. M., & Sher, K. J. (2005). Similarities and differences of longitudinal phenotypes across alternate indices of alcohol involvement: a methodologic comparison of trajectory approaches. *Psychology of Addictive Behaviors, 19*(4), 339.

Jenkins, J. M., Rasbash, J., & O'Connor, T. G. (2003). The role of the shared family context in differential parenting. *Developmental psychology, 39*(1), 99.

Jensen, P. S. (2003). Comorbidity and Child Psychopathology: Recommendations for the Next Decade. *Journal of Abnormal Child Psychology, 31*(3), 293-300.

Jensen, P. S., Hoagwood, K., & Zitner, L. (2006). What's in a name? Problems versus prospects in current diagnostic approaches. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology* (Vol. 1, pp. 24-40). New Jersey: John Wiley & Sons.

Jung, T., & Wickrama, K. A. S. (2008). An Introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Karami, H. (2012). An Introduction to Differential Item Functioning. *International Journal of Educational and Psychological Assessment, 11*, 59-76.

Kaslow, N. J., Deering, C. G., & Racusin, G. R. (1994). Depressed children and their families. *Clin Psychol Rev, 14*(1), 39-59.

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet Need for Mental Health Care Among U.S. Children: Variation by Ethnicity and Insurance Status. *The American Journal of Psychiatry, 159*(9), 1548-1555.

Katz, J. N., Larson, M. G., Phillips, C. B., Fossel, A. H., & Liang, M. H. (1992). Comparative Measurement Sensitivity of Short and Longer Health Status Instruments. *Medical Care, 30*(10), 917-925.

Kauer-Sant'Anna, M., Frey, B. N., Andreazza, A. C., Cereser, K. M., Gazalle, F. K., Tramontina, J., . . . Kapczinski, F. (2007). Anxiety comorbidity and quality of life in

bipolar disorder patients. *Canadian journal of psychiatry. Revue canadienne de psychiatrie, 52*(3), 175-181.

Kays, K., Gathercoal, K., & Buhrow, W. (2012). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior, 28*(1), 251-256.

Kazdin, A. E. (1993). Treatment of conduct disorder: Progress and directions in psychotherapy research. *Development and Psychopathology, 5*(1-2), 277-310.

Keiley, M., Bates, J., Dodge, K., & Pettit, G. (2000). A Cross-Domain Growth Analysis: Externalizing and Internalizing Behaviors During 8 Years of Childhood. *Journal of Abnormal Child Psychology, 28*(2), 161-179.

Kennedy, I. (2010). Getting it right for children and young people: Overcoming cultural barriers in the NHS so as to meet their needs. In D. o. Health (Ed.). London.

Kessler, R. C., Foster, C. L., Saunders, W. B., & Stang, P. E. (1995). Social consequences of psychiatric disorders, I: Educational attainment. *American Journal of Psychiatry, 152*(7), 1026-1032.

Kessler, R. C., McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., . . . Angermeyer, M. (2010). Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *The British Journal of Psychiatry, 197*(5), 378-385.

Kim, H. K., Capaldi, D. M., & Stoolmiller, M. (2003). Depressive symptoms across adolescence and young adulthood in men: Predictions from parental and contextual risk factors. *Development and Psychopathology, 15*(02), 469-495.

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: correspondence among child, parent and teacher reports. *Journal of child psychology and psychiatry, and allied disciplines, 34*, 991-1006.

Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011). New dimensions in the quantitative classification of mental illness. *Archives of general psychiatry, 68*(10), 1003-1011.

Kovacs, M., & Devlin, B. (1998). Internalizing Disorders in Childhood. *Journal Of Child Psychology And Psychiatry, 39*(1), 47-63.

Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The Structure and Stability of Common Mental Disorders (DSM-III-R): A Longitudinal-Epidemiological Study. . *107*(2), 216-227.

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is There a General Factor of Prevalent Psychopathology During Adulthood? *Journal of Abnormal Psychology, 121*(4), 971-977.

Lahey, B. B., Schwab-Stone, M., Goodman, S. H., Waldman, I. D., Canino, G., Rathouz, P. J., . . . Jensen, P. S. (2000). Age and gender differences in oppositional behavior and conduct problems: A cross-sectional household study of middle childhood and adolescence. *Journal of Abnormal Psychology, 109*(3), 488.

Lahey, B. B., Van Hulle, C. A., Singh, A. L., Waldman, I. D., & Rathouz, P. J. (2011). Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. *Archives of general psychiatry, 68*(2), 181-189.

Landgraf, J. M., Abetz, L., & Ware, J. E. (1999). *The CHQ: A User's Manual* Boston, MA: The Health Institute.

Legerstee, J. S., Verhulst, F. C., Robbers, S. C., Ormel, J., Oldehinkel, A. J., & van Oort, F. V. (2013). Gender-Specific Developmental Trajectories of Anxiety during Adolescence: Determinants and Outcomes. The TRAILS Study. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 22*(1), 26-34.

Lerner, R. M. (2006). Developmental Science, Developmental Systems, and Contemporary Theories of Human Development. In R. M. Lerner & W. R. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed., Vol. 1). Hoboken, NJ: John Wiley & Sons, Inc.

Lerner, R. M., Agans, J. P., DeSouza, L. M., & Gasca, S. (2013). Describing, Explaining, and Optimizing Within-Individual Change Across the Life Span: A Relational Developmental Systems Perspective. *Review of General Psychology, 17*(2), 179-183.

Leuven, E., & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing (Version revised 19 Jul 2012): Boston College Department of Economics.

Levitt, J. M., Saka, N., Romanelli, L. H., & Hoagwood, K. (2007). Early Identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*, 163-191.

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III—R disorders in high school students. *Journal of Abnormal Psychology, 102*(1), 133.

Lewis, M. (2000). Toward a development of psychopathology. In A. J. Sameroff, M. Lewis & S. M. Miller (Eds.), *Handbook of developmental psychopathology* (pp. 3-22). New York: Springer.

Lien, L., Tambs, K., Oppedal, B., Heyerdahl, S., & Bjertness, E. (2005). Is relatively young age within a school year a risk factor for mental health problems and poor school

performance? A population-based cross-sectional study of adolescents in Oslo, Norway. *BMC public health, 5*(1), 102-109.

Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions. *Journal of Abnormal Child Psychology, 31*(3), 285-291.

Liu, I. M., & Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics, 52*(4), 1223-1234.

Livingstone, S. (2008). Taking risky opportunities in youthful content creation: teenagers use of social networking sites for intimacy, privacy and self-expression. *New Media & Society, 10*(3), 393-411.

Lyon, G. R. (1996). Learning disabilities. *The future of children*, 54-76.

Macmillan, R., McMorris, B. J., & Kruttschnitt, C. (2004). Linked Lives: Stability and Change in Maternal Circumstances and Trajectories of Antisocial Behavior in Children. *Child Development, 75*(1), 205-220.

Magnusson, D. (1995). Individual Development: a holistic, integrated model. In M. Phyllis, H. E. Glen & L. Kurt (Eds.), *Examining lives in context : perspectives on the ecology of human development* Washington DC: American Psychological Association.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*(3), 391.

Martel, M. M. (2013). Sexual Selection and Sex Differences in the Prevalence of Childhood Externalizing and Adolescent Internalizing Disorders. *Psychological Bulletin, 139*(6), 1221-1259.

Martin, P., & de Francesco, D. (2014). *CAMHS payment systems and the prediction of treatment costs*. Paper presented at the CAMHS conference, Northampton.

Masten, A. S. (2001). Ordinary magic: Resilience processes in development. *American Psychologist, 56*(3), 227.

Masten, A. S., Burt, K. B., & Coatsworth, J. D. (2006). Competence and psychopathology in development. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology: risk, disorder and adaptation* (2nd ed., pp. 696-738). NewYork: Wiley and Sons.

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradovic, J., Riley, J. R., . . . Tellegen, A. (2005). Developmental cascades: linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental psychology, 41*, 733-746.

Maughan, B., Rowe, R., Messer, J., Goodman, R., & Meltzer, H. (2004). Conduct disorder and oppositional defiant disorder in a national sample: developmental epidemiology. *Journal Of Child Psychology And Psychiatry, 45*(3), 609-621.

Maughan, B., & Rutter, M. (2009). Development and Psychopathology: A Life Course Perspective *Rutter's Child and Adolescent Psychiatry* (pp. 160-181): Blackwell Publishing Ltd.

Mazerolle, P., Brame, R., Paternoster, R. A. Y., Piquero, A., & Dean, C. (2000). Onset Age, Persistance, and Offending Versatility: Comparisons Across Gender. *Criminology, 38*(4), 1143-1172.

McLeod, J. D., & Fettes, D. L. (2007). Trajectories of failure: The educational careers of children with mental health problems. *American Journal of Sociology, 113*(3), 653-701.

McNutt, L.-A., Wu, C., Xue, X., & Hafner, J. P. (2003). Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes. *American Journal of Epidemiology, 157*(10), 940-943.

Measelle, J. R., Stice, E., & Hogansen, J. M. (2006). Developmental Trajectories of Co-Occurring Depressive, Eating, Antisocial, and Substance Abuse Problems in Female Adolescents. *Journal of Abnormal Psychology 115*(3), 524-538.

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistant antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674-701.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology, 13*(02), 355-375.

Moffitt, T. E., Caspi, A., Dickson, N., Silva, P., & Stanton, W. (2009). Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and Psychopathology, 8*, 399.

Moffitt, T. E., Caspi, A., Harrington, H., Milne, B. J., Melchior, M., Goldberg, D., & Poulton, R. (2007). Generalized anxiety disorder and depression: childhood risk factors in a birth cohort followed to age 32. *Psychological Medicine, 37*(03), 441-452.

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. New York, NY US: Cambridge University Press.

Moilanen, K. L., Shaw, D. S., & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development and Psychopathology, 22*(3), 635-653.

Morin, A. J. S., Maïano, C., Nagengast, B., Marsh, H. W., Morizot, J., & Janosz, M. (2011). General Growth Mixture Analysis of Adolescents' Developmental Trajectories of Anxiety: The Impact of Untested Invariance Assumptions on Substantive Interpretations. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(4), 613-648.

Muris, P., Meesters, C., & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ). *European child & adolescent psychiatry, 12*(1), 1-8.

Muris, P., & Ollendick, T. H. (2005). The Role of Temperament in the Etiology of Child Psychopathology. *Clinical Child and Family Psychology Review, 8*(4), 271-289.

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Lopez, A. D. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990?2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet, 380*(9859), 2197-2223.

Murray, J., & Farrington, D. P. (2010). Risk factors for conduct disorder and delinquency: key findings from longitudinal studies. *Canadian journal of psychiatry. Revue canadienne de psychiatrie, 55*(10), 633-642.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Newbury Park, CA: Sage.

Muthén, B. O. (2006). The potential of growth mixture modelling. *Infant and Child Development, 15*(6), 623-625.

Muthén, B. O., & Khoo, S.-T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and individual differences, 10*(2), 73-101.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*(6), 882-891.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267-316.

Muthén, L. K., & Muthén, B. O. (2007). Mplus User's Guide (Sixth Edition). In M. Muthén (Ed.). Los Angeles, CA.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (Seventh Edition). Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (1999). Analyzing developmental trajectories: a semiparametric, group-based approach. *Psychological methods, 4*(2), 139-157.

Nagin, D. S., & Tremblay, R. E. (2005). Developmental Trajectory Groups: Fact or a Useful Statistical Fiction? *Criminology, 43*(4), 873-904.

National Institute for Mental Health in England. (2008). Mental Health Outcomes Compedium. London: Department of Health Publications.

Nazroo, J. Y. (1998). Genetic, Cultural or Socio-economic Vulnerability? Explaining Ethnic Inequalities in Health. *Sociology of Health & Illness, 20*(5), 710-730.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. A. Collins & J. L. Horn (Eds.), *Best Methods for the Analysis of Change* (pp. 92-106). Washington, DC: American Psychological Association.

Nikapota, A., & Rutter, M. (2009). Sociocultural/Ethnic Groups and Psychopathology. In M. Rutter & e. al. (Eds.), *Rutter's Child and Adolescent Psychiatry* (pp. 199-211): Blackwell Publishing Ltd.

Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: evidence and theory. *Psychological Bulletin, 101*(2), 259.

Nolen-Hoeksema, S., & Girgus, J. S. (1994). The Emergence of Gender Differences in Depression During Adolescence. . *Psychological Bulletin, 115*(3), 424-443.

Norcross, J. C., & Wampold, B. E. (2011). What works for whom: Tailoring psychotherapy to the person. *Journal of Clinical Psychology, 67*(2), 127-132.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling, 14*(4), 535-569.

O'Connor, T. G., McGuire, S., Reiss, D., Hetherington, E. M., & Plomin, R. (1998). Co-occurrence of depressive symptoms and antisocial behavior in adolescence: a common genetic liability. *Journal of Abnormal Psychology, 107*(1), 27.

Odgers, C. L., Caspi, A., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., . . . Moffitt, T. E. (2007). Prediction of differential adult health burden by conduct problem subtypes in males. *Archives of general psychiatry, 64*(4), 476-484.

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology, 20*(02), 673-716.

Olweus, D. (1993). *Bullying at school: what we know and what we can do*. Oxford: Blackwell Publishers.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Bloomsbury Publishing.

Pachter, L. M., Auinger, P., Palmer, R., & Weitzman, M. (2006). Do Parenting and the Home Environment, Maternal Depression, Neighborhood, and Chronic Poverty Affect Child Behavioral Problems Differently in Different Racial-Ethnic Groups? *Pediatrics, 117*(4), 1329-1338.

Paek, I., & Han, K. T. (2012). IRTPRO 2.1 for Windows (Item Response Theory for Patient-Reported Outcomes). *Applied Psychological Measurement*.

Panak, W. F., & Garber, J. (1992). Role of aggression, rejection, and attributions in the prediction of depression in children. *Development and Psychopathology, 4*(01), 145-165.

Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: a global public-health challenge. *The Lancet, 369*(9569), 1302-1313.

Patterson, C. J., Kupersmidt, J. B., & Vaden, N. A. (1990). Income Level, Gender, Ethnicity, and Household Composition as Predictors of Children's School-based Competence. *Child Development, 61*(2), 485-494.

Patterson, J., & Blum, R. W. (1996). Risk and resilience among children and youth with disabilities. *Archives of pediatrics & adolescent medicine, 150*(7), 692.

Pearl, J. (2009). *Causality: models, reasoning and inference* (Vol. 29). New York: Cambridge University Press.

Penfield, R. (2005). DIFAS: Differential Item Functioning Analysis System. *Applied Psychological Measurement,, 29*(2), 150-151.

Penfield, R. (2007). An approach for categorizing DIF in polytomous items. *Applied Measurement in Education, 20*(3), 335-355.

Penfield, R., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement, 43*(4), 295-312.

Petersen, M. A., Groenvold, M., Bjorner, J. B., Aaronson, N., Conroy, T., Cull, A., . . . Sullivan, M. (2003). Use of Differential Item Functioning Analysis to Assess the Equivalence of Translations of a Questionnaire. *Quality of Life Research, 12*(4), 373-385.

Piquero, A. R., Farrington, D. P., Nagin, D. S., & Moffitt, T. E. (2010). Trajectories of offending and their relation to life failure in late middle age: Findings from the Cambridge Study in Delinquent Development. *Journal of Research in Crime and Delinquency, 47*(2), 151-173.

Poulin, C., Hand, D., Boudreau, B., & Santor, D. (2005). Gender differences in the association between substance use and elevated depressive symptoms in a general adolescent population. *Addiction, 100*(4), 525-535.

Prensky, M. (2001). Digital Natives, Digital Immigrants: Part 1. *On the Horizon, 9*(5), 1-6.

Prior, M., Smart, D., Sanson, A., & Oberklaid, F. (1999). Relationships between learning difficulties and psychological problems in preadolescent children from a longitudinal sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(4), 429-436.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata. Third Edition*. College Station, TX: Stata Press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111-164.

Raine, A., Moffitt, T. E., Caspi, A., Loeber, R., Stouthamer-Loeber, M., & Lynam, D. (2005). Neurocognitive impairments in boys on the life-course persistent antisocial path. *Journal of Abnormal Psychology, 114*(1), 38.

Ramchandani, P., & Psychogiou, L. (2009). Paternal psychiatric disorders and children's psychosocial development. *Lancet, 374*(9690), 646-653.

Ravens-Sieberer, U., Auquier, P., Erhart, M., Gosch, A., Rajmil, L., Bruil, J., . . . the KIDSCREEN Group. (2007). The Kidscreen-27 quality of life measure for children and adolescent: psychometric results from a cross-cultural survey in 13 European countries. *Quality of Life Research, 16*(8), 1347-1356.

Ravens-Sieberer, U., Erhart, M., Rajmil, L., Herdman, M., Auquier, P., Bruil, J., . . . the European KIDSCREEN Group. (2010). Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. *Quality of Life Research, 19*(10), 1487-1500.

Ravens-Sieberer, U., Gosch, A., Rajmil, L., Erhart, M., Bruil, J., Power, M., . . . the European KIDSCREEN Group. (2007). The KIDSCREEN-52 Quality of Life measure for children and adolescents: Psychometric results from a cross-cultural survey in 13 European Countries. *Value in Health, 11*(4), 645–658.

Ray, J. J. (1983). Reviving the problem of acquiescent response bias. *Journal of Social Psychology, 121*, 81-96.

Raykov, T., Marcolides, G. A. (2011). Introduction to psychometric theory. New York, NY: Routledge.

Reinherz, H. Z., Giaconia, R. M., Lefkowitz, E. S., Pakiz, B., & Frost, A. K. (1993). Prevalence of Psychiatric Disorders in a Community Population of Older Adolescents.

Journal of the American Academy of Child and Adolescent Psychiatry, 32(2), 369-377.

Renouf, A. G., Kovacs, M., & Mukerji, P. (1997). Relationship of depressive, conduct, and comorbid disorders and social functioning in childhood. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(7), 998-1004.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children manual*. Circle Pines, MN: American Guidance Service.

Riglin, L., Frederickson, N., Shelton, K. H., & Rice, F. (2013). A longitudinal study of psychological functioning and academic attainment at the transition to secondary school. *Journal of adolescence, 36*(3), 507-517.

Riglin, L., Petrides, K. V., Frederickson, N., & Rice, F. (2014). The relationship between emotional problems and subsequent school attainment: A meta-analysis. *Journal of adolescence, 37*(4), 335-346.

Ritter, P., Lorig, K., Laurent, D., & Matthews, K. (2004). Internet versus mailed questionnaires: a randomized comparison. *Journal of Medical Internet Research, 6*(3).

Roberts, N., & Cawthorpe, D. (1995). Immigrant child and adolescent psychiatric referrals: a five-year retrospective study of Asian and Caucasian families. *Canadian journal of psychiatry. Revue canadienne de psychiatrie*, *40*(5), 252-256.

Rogers, H. J. (2005). Differential Item Functioning *Encyclopedia of Statistics in Behavioral Science*: John Wiley & Sons, Ltd.

Roisman, G., Monahan, K., Campbell, S., Steinberg, L., & Cauffman, E. (2010). Is adolescence-onset antisocial behavior developmentally normative? *Development And Psychopathology, 22*(2), 295-311.

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: replication and extension using a nationwide Canadian survey. *Developmental psychology, 46*(5), 995.

Rosen, F. S. (2000). A brief history of immunodeficiency disease. *Immunological Reviews, 178*(1), 8-12.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician, 39*(1), 33-38.

Roza, S., Hofstra, M., van der Ende, J., & Verhulst, F. C. (2003). Stable prediction of mood and anxiety disorders based on behavioral and emotional problems in childhood: A 14-year follow-up during childhood, adolescence, and young adulthood. *American Journal of Psychiatry, 160*(12), 2116–2121.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (1998). Gender development. *Handbook of child psychology*.

Rushton, J., Forcier, M., & Schectman, R. (2002). Epidemiology of depressive symptoms in the national longitudinal study of adolescent health. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(2), 199-205.

Rust, J., & Golombok, S. (2009). *Modern psychometrics : the science of psychological assessment* (3rd ed.). London: Routledge.

Rutter, M. (2007). Proceeding From Observed Correlation to Causal Inference: The Use of Natural Experiments. *Perspectives on Psychological Science, 2*(4), 377-395.

Rutter, M., Champion, L., Quinton, D., Maughan, B., & Pickles, A. (1995). Understanding individual differences in environmental-risk exposure. In P. Moen, G. H. Elder & K. Luscher (Eds.), *Examining lives in context: Perspectives on the ecology of human development* (pp. 61-96). Washington DC: American Psychological Association.

Sadowski, H., Ugarte, B., Kolvin, I., Kaplan, C., & Barnes, J. (1999). Early life family disadvantages and major depression in adulthood. *The British Journal of Psychiatry, 174*(2), 112-120.

Samejima, F. (1997). Graded response model. In W. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* N.Y.: Springer.

Santor, D. A., Ramsay, J. O., & Zuroff, D. C. (1994). Non-Parametric Item Analysis of the Beck Depression Inventory: Evaluating Gender Item Bias and Response Option Weights. *Psychological Assessment, 6*(3), 255-270.

Schmidt, L. J., Garratt, A. M., & Fitzpatrick, R. (2000). Instruments for Mental Health: A review. Oxford: Patient-Reported Health Instruments Group.

Schmidt, L. J., Garratt, A. M., & Fitzpatrick, R. (2001). Instruments for Children and Adolescents: a Review. Oxford: Patient-Reported Health Instruments Group

Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing Health Status and Quality of Life instruments: Attributes and Review Criteria. *Quality of Life Research, 11,* 193-205.

Sharma, S. (1970). Manifest anxiety and school achievement of adolescents. *Journal of Consulting and Clinical Psychology, 34*(3), 403.

Sharp, C., Goodyer, I. M., & Croudace, T. (2006). The Short Mood and Feelings Questionnaire (SMFQ): a unidimensional item response theory and categorical data factor analysis of self-report ratings from a community sample of 7-through 11-year-old children. *Journal of Abnormal Child Psychology, 34*(3), 379-391.

Shoshani, A., & Steinmetz, S. (2013). Positive Psychology at School: A School-Based Intervention to Promote Adolescents' Mental Health and Well-Being. *Journal of Happiness Studies*, 1-23.

Silberg, J., Rutter, M., D'Onofrio, B., & Eaves, L. (2003). Genetic and environmental risk factors in adolescent substance use. *Journal of Child Psychology and Psychiatry, 44*(5), 664-676.

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric Disorders in Children With Autism Spectrum Disorders: Prevalence, Comorbidity, and Associated Factors in a Population-Derived Sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(8), 921-929.

Smith, D. (2005). Explaining ethnic variations in crime and antisocial behavior in the United Kingdom. In M. Rutter & M. Tienda (Eds.), *Ethnicity and causal mechanisms* (pp. 174-203). New York: Cambridge University Press.

Solans, M., Pane, S., Estrada, M.-D., Serra-Sutton, V., Berra, S., Herdman, M., . . . Rajmil, L. (2008). Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 11*, 742-764.

Sroufe, L. A. (2007). The place of development in developmental psychopathology. In A. S. Masten (Ed.), *Multilevel dynamics in developmental psychopathology: The Minnesota Symposia on Child Psychology* (Vol. 34, pp. 285-299). USA: Lawrence Erlbaum Associates.

Sroufe, L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development, 55*, 17-29.

StataCorp. (2011). Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.

Sterba, K., Prinstein, J., & Cox, J. (2007). Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences. *Development and Psychopathology, 19*(2), 345-366.

Suhrcke, M., Pillas, D., & Selai, C. (2008). Economic aspects of mental health in children and adolescents *Social cohesion for mental well-being among adolescents. .* Copenhagen: WHO Regional Office for Europe.

Suler, J. (2006). The online disinhibition effect. *Cyber Psychology and Behavior, 7*(3), 321-333.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42.

Thelen, E., & Smith, L. B. (2006). Dynamic Systems Theories. In R. M. Lerner & W. R. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (Sixth ed., Vol. 1). Hoboken, NJ: Wiley.

Thompson, A. H., Barnsley, R. H., & Battle, J. (2004). The relative age effect and the development of self-esteem. *Educational Research, 46*(3), 313-320.

Turner, C. F., Ku, L., Rogers, S. M., Lindberg, L. D., Pleck, J. H., & Sonenstein, F. L. (1998). Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting with Computer Survey Technology. *Science, 280*(5365), 867-873.

Twenge, J. M., & Nolen-Hoeksema, S. (2002). Age, gender, race, socioeconomic status, and birth cohort difference on the children's depression inventory: A meta-analysis. *Journal of Abnormal Psychology, 111*(4), 578.

UN Convention on the Rights of the Child, United Nations General Assembly (1989).

Vaillancourt, T., Brittain, H. L., McDougall, P., & Duku, E. (2013). Longitudinal links between childhood peer victimization, internalizing and externalizing problems, and academic functioning: Developmental cascades. *Journal of Abnormal Child Psychology, 41*(8), 1203-1215.

Van Cleave, J., & Davis, M. M. (2006). Bullying and peer victimization among children with special health care needs. *Pediatrics, 118*(4), e1212-e1219.

van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.

van Lier, P. A. C., Vitaro, F., Barker, E. D., Brendgen, M., Tremblay, R. E., & Boivin, M. (2012). Peer Victimization, Poor Academic Achievement, and the Link Between Childhood Externalizing and Internalizing Problems. *Child Development, 83*(5), 1775-1788.

Varese, F., Smeets, F., Drukker, M., Lieverse, R., Lataster, T., Viechtbauer, W., . . . Bentall, R. P. (2012). Childhood adversities increase the risk of psychosis: a meta-analysis of patient-control, prospective-and cross-sectional cohort studies. *Schizophrenia bulletin, 38*(4), 661-671.

Verhulst, F. C., & Koot, H. M. (1992). *Child psychiatric epidemiology: Concepts, methods, and findings*. Thousand Oaks:CA: Sage Publications, Inc.

Verhulst, F. C., & van der Ende, J. (2006). *Assessment Scales in Child and Adolescent Psychiatry*. Taylor & Francis.

Verhulst, F. C., & Van der Ende, J. (2009). Using Rating Scales in a Clinical Context. *Rutter's Child and Adolescent Psychiatry* (pp. 289-298). Blackwell Publishing Ltd.

von Eye, A. (2010). Developing the person-oriented approach: theory and methods of analysis. *Development and Psychopathology, 22*(2), 277.

von Eye, A., & Bergman, L. (2003). Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. *Development and Psychopathology, 15*, 553–580.

Wade, T. J., Cairney, J., & Pevalin, D. J. (2002). Emergence of gender differences in depression during adolescence: National panel results from three countries. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(2), 190-198.

Walker, C. M. (2011). What's the DIF? Why Differential Item Functioning Analyses Are an Important Part of Instrument Development and Validation. *Journal of Psychoeducational Assessment, 29*(4), 364-376.

Wampold, B. E. (2013). *The great psychotherapy debate: Models, methods, and findings*. Mahwah, NJ: Erlbaum.

Wang, M., & Bodner, T. E. (2007). Growth Mixture Modeling: Identifying and Predicting Unobserved Subpopulations With Longitudinal Data. *Organizational Research Methods, 10*(4), 635-656.

Watson, D., & Clark, L. A. (1984). Negative affectivity: the disposition to experience aversive emotional states. *Psychological Bulletin, 96*(3), 465.

Weeks, M., Coplan, R. J., & Kingsbury, A. (2009). The correlates and consequences of early appearing social anxiety in young children. *Journal of anxiety disorders, 23*(7), 965-972.

Weiss, B., Süsser, K., & Catron, T. (1998). Common and specific features of childhood psychopathology. *Journal of Abnormal Psychology, 107*(1), 118.

Werry, J. S., Reeves, J. C., & Elkind, G. S. (1987). Attention deficit, conduct, oppositional, and anxiety disorders in children: I. A review of research on differentiating characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry, 26*(2), 133-143.

WHO. (2008). The global burden of disease: 2004 update. Switzerland: World Health Organization.

Whitehead, L. (2011). Methodological issues in Internet-mediated research: a randomized comparison of internet versus mailed questionnaires. *Journal of medical Internet research, 13*(4).

Wijndaele, K., Matton, L., Duvigneaud, N., Lefevre, J., Duquet, W., Thomis, M., . . . Philippaerts, R. (2007). Reliability, equivalence and respondent preference of computerized versus paper-and-pencil mental health questionnaires. *Computers in human behavior, 23*(4), 1958-1970.

Wolpert, M., Aitken, J., Syrad, H., Munroe, M., Saddington, C., Trustam, E., . . . Croudace, T. (2008). Review and recommendations for national policy for England for the use of mental health outcome measures with children and young people. London: Department for Children, Schools and Families.

Wolpert, M., Deighton, J., Patalay, P., Martin, A., Fitzgerald-Yau, N., Demir, E., . . . Frederikson, N. (2011). Me and my school: Findings from the National Evaluation of Targeted Mental Health in Schools. Nottingham: DFE.

Wright, A. G., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. . *Journal of Abnormal Psychology, 122*(1), 281-294.

Yammarino, F. J., & Atwater, L. E. (1993). Understanding self-perception accuracy: Implications for human resource management. *Human Resource Management, 32*(2), 231-247.

Yang, C.-C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50*(4), 1090-1104.

Yung, Y.-F., Thissen, D., & McLeod, L. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113-128.

Zachar, P., & Kendler, K. (2007). Psychiatric disorders: a conceptual taxonomy. *American Journal of Psychiatry, 164*(4), 557-565.

Zahn-Waxler, C., Robinson, J. L., & Emde, R. N. (1992). The development of empathy in twins. *Developmental psychology, 28*(6), 1038.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(4), 561-577.

Appendices

Appendix A: Systematic review of measures (update 2008-2012)

Systematic review of outcome measures in child and adolescent mental health: 2008-2012

This review was carried out as an extension to an existing review, (Wolpert et al., 2008)[1] to identify new measures that might have come out recently. The original systematic review identified 12 measures after five stages which met a range of inclusion and exclusion criteria and had a broad range of psychometric evidence supporting them.

The full Wolpert et al., (2008) report can be found here: http://www.ucl.ac.uk/clinical-psychology/EBPU/publications/reports.php

List of measures identified by Wolpert et al. (2008)

Achenbach System of Empirically Based Assessment (ASEBA)
Beck Youth Inventories (BYI)
Behaviour Assessment System for Children (BASC)
Behavioural and Emotional Rating Scale (BERS)
Child Health Questionnaire (CHQ)
Child Symptom Inventories (CSI)
Health of the National Outcome Scale for Children and Adolescents (HoNOSCA)

[1] Wolpert M, Aitken J, Syrad H, Munroe M, Saddington C, Trustam E, et al. (2008) Review and recommendations for national policy for England for the use of mental health outcome measures with children and young people. London: Department for Children, Schools and Families.

Retrieved from:

http://www.ucl.ac.uk/clinical-psychology/EBPU/publications/reports.php

Kidscreen
Pediatric Symptom Checklist (PSC)
PedsQL Present Functioning (PedsQL)
Strengths and Difficulties Questionnaire (SDQ)
Youth Outcome Questionnaire (YOQ)

The methodology used in the current report is the same as in the initial review to ensure parity and same rigour of method. This review update was carried out in early 2012 (Feb-Mar) and involved four stages

Stage 1: Setting the parameters of the review

Stage 2: Sorting the Results

Stage 3: Inclusion and Exclusion Criteria

Stage 4: Details of measures found

Stage 1: Setting parameters of the review

Child mental health outcome measures for the purposes of this review were defined as any questionnaire, measure or approach to measurement that seeks to provide measurement of mental health in children and young people (up to 18 years). Search terms were developed to capture these defining features by a) splitting the terms child mental health outcomes measure into three defining features: I) 'measurement', II) 'mental health' and III) 'child' and b) generating a list of words or phrases that reflect each of these features (see table A below).

Table A:

Factor	Related Terms
1) Measures and approaches to measurement	Measure; questionnaire; survey; checklist; check list; tool; rating scale; scale; repository
2) Mental health and psychological well-being	Mental health; quality of life; psychological adjustment; behaviour problems; emotional problems; mental illness; mental disorder; psychiatric disorder; behavioural and emotional difficulties; social difficulties; social and behavioural difficulties; conduct problems; internalising; externalising; depressive symptoms; antisocial; self-esteem; pride; prosocial behaviour; sense of belonging; hopefulness; well being; positive self-regard; aggression; anxiety; depression; mood; feeling
3) Children	Children; adolescents; paediatrics

Search of key databases

Initial searches focused on 4 key databases: MEDLINE, EMBASE, PsychInfo, ERIC by conducting searches in each database for the terms relating to 1) measures and approaches to measurement, 2) mental health and psychological well-being and 3) children.

Searches for these terms were then conducted in the thesaurus facility within each of the four databases. These thesaurus searches were carried out to identify subject heading or MeSH terms under which papers of interest would be catalogued in each database. Searches in thesauruses of each database identify terms that map onto MeSH headings or subject headings relevant to child mental health outcome measures which are specific to each database. Subject headings and MeSH terms for each database were then refined to discard terms that were too broad or too narrow to capture our search criteria. These final MeSH headings/subject headings were used to run searches in each of the respective databases in order to identify relevant papers (see Table B below).

As this review is an update, limits were set to restrict results to just those that were 2008 and after (as the previous review was carried out in early 2008).

Searches that resulted in over 100 hits were first subject to basic filtering to discard obviously irrelevant hits. This involved the following exclusion criteria being applied to the title:

- that the paper was not related to children's mental health outcome measures
- that the paper was not in English

Table B: the search terms used in the different databases and the number of results

Database	Search terms for subject heading and mesh headings	Initial Hits*	Hits After Basic Filtering *
EMBASE	(Child or children or adolescent) in SU and (child psychiatry or child psychology or mental disease or adjustment disorder or behaviour disorder or emotional disorder or mental instability or mood disorder) in SU)and((psychological test or named inventories or questionnaires or rating scales) in SU) and (LA:EMBV = ENGLISH)	1	
ERIC	(Check-lists OR Measurement OR Questionnaires OR Rating-Scales).DE. AND (Mental-health OR Behaviour-problems OR Emotional-Adjustment OR Emotional-Disturbances OR Emotional-Problems OR Mental-Disorders OR Psychopathology).DE. AND (Children OR Adolescents).DE. AND LG=ENGLISH	410	40
Medline	((mental health or Mental Disorders or emotional problems or Emotional Disturbances or Behaviour Disorders or adaptation psychological or Psychological adjustment) in MJME)and((child or adolescent) in MJME)and((Questionnaires or psychiatric rating scales or treatment outcome) in MJME)	448	5
Psych Info	(Questionnaires or measurement or surveys or rating scales or Lickert scales or symptom checklists) in SU)and((child psychopathology or child psychiatry or child psychology or adolescent psychopathology or adolescent psychiatry or adolescent psychology) in SU)and((Mental health or community mental health or well being or emotional adjustment or social adjustment or mental disorders or adjustment disorders or psychological stress or behaviour problems or internalization or externalization or psychiatric symptoms or distress) in SU)	8	

*NB 'hits' refers to the number of papers identified in each database

Stage 2: Sorting the Results

In stage 2 the results were further sorted based on more specific search criteria by assessing the abstracts of the papers. This involved applying the criteria listed below (see Box A) to the hits identified in Stage 1 after filtering.

Box A

Hit excluded if:

1) no child mental health outcome measure was mentioned in abstract
2) the measure mentioned was too narrow to provide a broad assessment of mental health (e.g., focused exclusively on just personality disorders or just schizophrenia)
3) the paper referred to a measure not used with children
4) the paper was not in English
5) the paper was a duplicate of a previous hit within the database
6) the paper referred to an assessment or DSM diagnosis

At the end of this stage 20 measures were identified which are listed below

1. Strengths and Difficulties Questionnaire (SDQ)
2. Achenbach System of Empirically Based Assessment (ASEBA); Child Behaviour Checklist (CBCL; different versions)
3. KiGGS(German Health Interview & examination survey for children and adolescents)
4. Brief Problem Checklist (part of the ASEBA)
5. Beck's Depression Inventory (BDI)
6. Diagnostic Infant & Pre school assessment
7. Behaviour Problems Inventory
8. Nisonger child behaviour rating scale
9. Diagnostic Interview for Children and Adolescents (DICA-R)
10. Pre-School and Kindergarten Behaviour Scales (PKBS-20)
11. Behavioural and Emotional Rating Scale BERS-2
12. Questionnaire Based on HEADSS Approach QBH-16
13. Brief Child and Family Phone Interview (BCFPI)

14. Ontario Child Health Scale – Revised (OCHS-R)
15. Connors rating scales
16. Systematic screening for behaviour disorders
17. Disruptive behaviour rating scale
18. Preschool behavioural and emotional rating scale
19. Social Skills Improvement System (SSIS) rating scales
20. Child Health and Illness Profile (CHIPS)

Stage 3: Inclusion and Exclusion Criteria

Inclusion and exclusion criteria that were used to filter and select measures in the original review were used in this stage.

An additional exclusion criteria was added, measures that were included in the previous review (at the end of stage 3, N=43) were excluded at this stage as the aim of this review was to identify measures that were recent.

Inclusion and exclusion criteria for measures or approaches to measurement

Inclusion criteria:

To include any questionnaire, measure or approach to outcome evaluation

1. that seeks to provide measurement of generic mental health in children and young people (up to 18)
2. that is either multi-dimensional or uni-dimensional
3. that can be completed by child or parents/carers with the possible addition of professionals
4. that has been validated in a child or adolescent context, even if not originally developed for this purpose
5. that is available in English language

6. that can be used with a reasonably wide age range (e.g., not just for preschoolers)

Exclusion criteria:

To exclude questionnaires, measures or approaches to outcome evaluation

1. that were identified in the original review.
2. that are not available in English
3. that do not measure mental health outcomes
4. that do not cover broad range of difficulties i.e. concern only specific mental disorders or domains e.g., ADHD, schizophrenia, physical problems, eating disorders, self-harm, OCD, psychosis, autism, specific learning difficulties, phobias etc, or internalizing or externalizing only
5. that are not used with children
6. that are based on professional report only (e.g., teacher or clinician);
7. that take over 30 minutes to complete
8. that provide open-ended responses that have to be manually coded
9. where the age range was too narrow (e.g., pre-school version of the measure only)
10. have not been used with a variety of populations (i.e. only used with very specialist groups)

Application of the above criteria resulted in a reduced list of 4 measures. Those measures where there was not enough information yet to judge whether they met the inclusion/exclusion criteria remained included at this stage.

Table C Showing 4 measures identified after Stage 3

Measure
Brief Problem checklist (BPC)
Diagnostic Infant & Pre school assessment
Questionnaire Based on HEADSS Approach (QBH-16)
Brief Child and Family Phone Interview (BCFPI)

Stage 4: Detailed exploration of measures identified in Stage 3

The 4 measures identified after stage 3 were explored in greater detail by finding papers and looking at their websites and are briefly explained below.

1. Brief Problem Checklist (BPC)

   The Brief Problem checklist is a 12-item measure. It has child and caregiver versions that were created by applying Item Response Theory and Factor Analysis to the Youth Self Report and Child Behaviour Checklist (of the ASEBA) respectively. There is one internalizing factor and externalising factor. (Chorpita, Reise, Weisz, Grubbs, Becker & Krull, 2010)

2. Diagnostic infant and pre-school assessment (DIPA)

   The DIPA is an interview and was designed to diagnose children aged 1- 6 years with the mention that it can be used with children upto age 8 if required. It measures 13 disorders based on DSM-IV criteria. It consists of 517 questions that require responses and is 47 pages long (Manual, 2008; Sheeringa & Haslett, 2010). Based on this we can exclude it from our final measures as it doesn't meet the inclusion criteria of needing to be less than 30 minutes long.

3. QBH-16

   The Questionnaire based on the HEADSS (Home, education, activities, drugs, sex and suicide) approach is a 16 item questionnaire with 2 items from each HEADSS

domain and 4 additional items regarding mental disorder or risk behaviours. It is aimed to be a general screening measure for adolescents aged 12-17 years. Items 1-6 completed by parent and items 7-16 by the adolescent and a total score is derived. Average completing time is 20 minutes (Hagel, Mainieri, Zeni & Wagner, 2009).

4. Brief Child and Family Phone Interview

The BCFPI is described as a computer assisted clinical intake and outcomes interview. It is designed for children and adolescents aged 3-18 years and covers broad mental health outcomes such as emotional and behavioural problems and also asks questions regarding child functioning, family functioning, risk and protective factors and records presence of 16 more specific, less common conditions such as eating disorders, fears, compulsions etc. (BCFPI website). It takes 30-45 minutes (excluded as it does not meet the inclusion criteria of being under 30 min) (Cunningham, Boyle, Hong, Pettingill & Bohaychuk, 2009; BCFPI website).

Conclusion

This review update identified two newer measures (BPC, Chorpita et al., 2010; QBH-16, Hagel et al., 2009) that meet the criteria set out by the review. Compared to the measures identified by the original review (Wolpert et al., 2008) which are more widely used and have lots of psychometric evidence of their properties both measures identified in this update are in early days of development and do not yet have sufficient psychometric evidence to fully support their use more widely.

References (for Appendix A)

Chorpita, B. F., Reise, S., Weisz, J. R., Grubbs, K., Becker, K. D., & Krull, J. L. (2010). Evaluation of the brief problem checklist: Child and caregiver interviews to measure clinical progress. Journal of Consulting and Clinical Psychology, 78(4), 526-536.

Cunningham, C. E., Boyle, M. H., Hong, S., Pettingill, P., & Bohaychuk, D. (2009). The brief child and family phone interview (BCFPI): 1. rationale, development, and description of a computerized children's mental health intake and outcome assessment tool. Journal of Child Psychology and Psychiatry, 50(4), 416-423. DOI:10.1111/j.1469-7610.2008.01970.x

Hagel, L. D., Mainieri, A. S., Zeni, C. P., & Wagner, M. B. (2009). Brief report: Accuracy of a 16-item questionnaire based on the HEADSS approach (QBH-16) in the screening of mental disorders in adolescents with behavioral problems in secondary care. Journal of Adolescence, 32(3), 753-761. DOI: /10.1016/j.adolescence.2008.08.009

Scheeringa, M. S., & Haslett, N. (2010). The reliability and criterion validity of the diagnostic infant and preschool assessment: A new diagnostic instrument for young children. Child Psychiatry and Human Development, 41(3), 299-312.

Wolpert M, Aitken J, Syrad H, Munroe M, Saddington C, Trustam E, et al. (2008) Review and recommendations for national policy for England for the use of mental health outcome measures with children and young people. London: Department for Children, Schools and Families.Retrieved from: http://www.ucl.ac.uk/clinical-psychology/EBPU/publications/reports.php

Appendix B: M&MS questionnaire

The following are the items in the *emotional difficulties* and *behavioural difficulties* scales of the M&MS questionnaire:

Emotional difficulties

I feel lonely

I cry a lot

I am unhappy

Nobody likes me

I worry a lot

I have problems sleeping

I wake up in the night

I am shy

I feel scared

I worry when I am at school

Behavioural difficulties

I get very angry

I lose my temper

I hit out when I am angry

I do things to hurt people

I am calm

I break things on purpose

Appendix C: Categorisation of diagnosis and presenting problems

Classifying system of the diagnosis and presenting problems based on two clinicians independent ratings

KEY: ED= emotional difficulty, BD= behavioural difficulty, EBD= Emotional and behavioural difficulties, ADHD= Hyperactivity and OTH= other

Diagnoses	Grouping Code
1. Adjustment disorder	1. EBD
2. Anxiety disorder	2. ED
3. Autism Spectrum Disorder (ASD)	3. OTH
4. Conduct disorder	4. BD
5. Depression	5. ED
6. Elective mutism	6. ED
7. Hyperkinetic Disorder (ADHD)	7. ADHD
8. Learning Difficulties	8. OTH
9. Low mood	9. ED
10. Mild depression	10. ED
11. Mild Mental retardation	11. OTH
12. Neurasthenia	12. OTH
13. Non-organic Insomnia	13. ED
14. OCD	14. ED
15. ODD	15. BD
16. PTSD	16. ED
17. Phobic Anxiety disorder	17. ED
18. Reactive attachment disorder	18. OTH
19. Self-harm	19. ED
20. Separation anxiety	20. ED
21. Tic disorder	21. OTH
22. Tourette's	22. OTH
23. Trichotillomania	23. ED
24. Unspecified behavioural and emotional problems (F98.9)	24. EBD

Presenting Problems	Grouping Code
1. Aggression	1. BD
2. Anger	2. EBD
3. Anxiety	3. ED
4. Anxiety eating after choking, weight loss	4. ED
5. Attachment problems	5. ED
6. Behavioural difficulties	6. BD
7. Being bullied	7. OTH
8. Concentration problems	8. ADHD
9. Conduct problems	9. BD
10. Delusions	10. OTH
11. Depression	11. ED
12. Difficult behaviour	12. BD
13. Disability	13. OTH
14. Disruptive behaviours	14. BD
15. Emotional difficulties	15. ED
16. Emotional problems	16. ED
17. Generalised anxiety	17. ED
18. Hallucinations	18. OTH
19. Hyperactivity	19. ADHD
20. Learning Difficulties	20. OTH
21. Low mood	21. ED
22. Low self-esteem	22. ED
23. Mood disorder	23. ED
24. Not sleeping	24. ED
25. Panic	25. ED
26. Peer relationships difficulties	26. OTH
27. Self-harm	27. ED
28. Tics	28. OTH
29. Trauma symptoms	29. ED
30. Weight loss	30. OTH